

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 7, 2002, 10:33:38 : Search time 20.04 Seconds  
(without alignments)  
1784.385 Million cell updates/sec

Title: US-09-842-256-2

Perfect score: 7631  
Sequence: 1 MSGMGENTSPDSRAETRRKK.....NQLPGMDIKEGDTRKCYC 1464

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 231628 seqs, 2442594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*

- 1: /cgn2-6/ptodata/2/1aa/5A\_COMB.pep.\*
- 2: /cgn2-6/ptodata/2/1aa/5B\_COMB.pep.\*
- 3: /cgn2-6/ptodata/2/1aa/6A\_COMB.pep.\*
- 4: /cgn2-6/ptodata/2/1aa/6B\_COMB.pep.\*
- 5: /cgn2-6/ptodata/2/1aa/PTUS\_COMB.pep.\*
- 6: /cgn2-6/ptodata/2/1aa/Backfile1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7631	100.0	1464	US-08-891-640-2	Sequence 2, Appli
2	1037.5	13.6	1036	US-08-891-640-3	Sequence 3, Appli
3	317	4.2	846	US-08-885-291-55	Sequence 55, Appli
4	317	4.2	846	US-09-107-847-2	Sequence 55, Appli
5	317	4.2	846	US-09-496-672-55	Sequence 55, Appli
6	309.5	4.1	855	US-08-816-693A-2	Sequence 2, Appli
7	309.5	4.1	855	US-08-885-291-2	Sequence 2, Appli
8	309.5	4.1	855	US-09-496-672-2	Sequence 2, Appli
9	292	3.8	870	US-08-785-241-4	Sequence 2, Appli
10	289	3.8	816	US-08-785-310A-8	Sequence 4, Appli
11	288	3.8	816	US-08-816-693A-53	Sequence 53, Appli
12	288	3.8	816	US-08-885-291-53	Sequence 53, Appli
13	288	3.8	816	US-09-496-672-53	Sequence 53, Appli
14	286.5	3.7	875	US-08-971-188-8	Sequence 8, Appli
15	283.5	3.6	875	US-08-785-241-5	Sequence 5, Appli
16	277	3.6	2414	US-08-227-536-2	Sequence 2, Appli
17	277	3.6	2414	PCT-US95-04682-2	Sequence 2, Appli
18	269.5	3.5	826	US-08-785-241-6	Sequence 2, Appli
19	269.5	3.5	826	US-08-480-473B-2	Sequence 6, Appli
20	269.5	3.5	826	US-08-915-213-2	Sequence 2, Appli
21	269.5	3.5	826	US-09-148-547-2	Sequence 2, Appli
22	269.5	3.5	826	US-09-235-217-2	Sequence 2, Appli
23	269.5	3.5	826	PCT-US96-10251-2	Sequence 2, Appli
24	269	3.5	626	US-08-971-188-10	Sequence 10, Appli
25	267	3.5	747	US-08-816-693A-51	Sequence 51, Appli
26	267	3.5	747	US-08-885-291-51	Sequence 51, Appli
27	267	3.5	747	US-09-496-672-51	Sequence 51, Appli

28	265	3.5	824	3	US-08-885-291-52	Sequence 52, Appli
29	265	3.5	824	4	US-09-496-672-52	Sequence 52, Appli
30	262	3.4	824	2	US-08-816-693A-52	Sequence 52, Appli
31	259	3.4	824	2	US-08-785-310A-7	Sequence 7, Appli
32	258.5	3.4	3969	4	US-08-061-376-5	Sequence 5, Appli
33	252	3.3	2441	1	US-08-194-468-2	Sequence 2, Appli
34	252	3.3	2441	3	US-08-961-739-2	Sequence 2, Appli
35	248	3.2	805	2	US-08-480-473B-4	Sequence 4, Appli
36	248	3.2	805	3	US-08-915-213-4	Sequence 4, Appli
37	248	3.2	805	4	US-09-235-217-4	Sequence 4, Appli
38	248	3.2	805	5	PCT-US96-10251-4	Sequence 4, Appli
39	245.5	3.2	810	1	US-08-785-241-7	Sequence 7, Appli
40	235	3.1	848	1	US-08-045-806-4	Sequence 4, Appli
41	235	3.1	848	1	US-08-366-051B-4	Sequence 4, Appli
42	234.5	3.1	1187	1	US-08-320-559-28	Sequence 28, Appli
43	234.5	3.1	1187	3	US-08-545-860D-28	Sequence 28, Appli
44	234.5	3.1	1187	5	PCT-US94-04496-28	Sequence 28, Appli
45	234.5	3.1	1210	1	US-08-320-559-26	Sequence 26, Appli

## ALIGNMENTS

RESULT 1  
US-08-891-640-2  
Sequence 2, Application US/08891640  
Patent No. 6268173  
GENERAL INFORMATION:  
APPLICANT: Chamdon, Pierre  
APPLICANT: Gronemeyer, Hinrich  
APPLICANT: Voegel, Johannes  
APPLICANT: Iutz, Yves  
TITLE OF INVENTION: Transcriptional Intermediary Factor-2  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESS: Sterne, Kessler, Goldstein & Fox P.L.L.C.  
STREET: 1100 New York Avenue, NW, Suite 600  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20005-3934  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/891,640  
FILING DATE: Herewith  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/021,247  
FILING DATE: 12-JUL-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Steffe, Eric K.  
REGISTRATION NUMBER: 36,688  
REFERENCE/DOCKET NUMBER: 1383.0130001/EKS  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-371-2600  
TELEFAX: 202-371-2540  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1464 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-891-640-2

Query Match 100.0%; Score 7631; DB 4; Length 1464;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1464; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MSGMGENTSDPSAEIRKRECEPDOLGSPKRNTEKRNKEQNKYIEBELIIFANFNDI 60
Db 1 MSGMGENTSDPSAEIRKRECEPDOLGSPKRNTEKRNKEQNKYIEBELIIFANFNDI 60
QY 61 DNENFPDKCALIKETVKOIROIKEOKAAANIDEVOKSDVSTGGVDKALGPMML 120
Db 61 DNENFPDKCALIKETVKOIROIKEOKAAANIDEVOKSDVSTGGVDKALGPMML 120
QY 121 EALDGEFFVNLNENGVSENVTOYLRYNOEELMKNKSVSILHVGDBTEFVNLKPKSI 180
Db 121 EALDGEFFVNLNENGVSENVTOYLRYNOEELMKNKSVSILHVGDBTEFVNLKPKSI 180
QY 181 VNGSGSGBEPBRNRSHTEFNCRLVKPLPDSEEGHNOEAKOYETMOCFAVQPKSIKE 240
Db 181 VNGSGSGBEPBRNRSHTEFNCRLVKPLPDSEEGHNOEAKOYETMOCFAVQPKSIKE 240
QY 241 EGDLDOSCLICVARVPKMRPLPSESEFTTODLOGKITSLDSTMRAMKPGWEDLV 300
Db 241 EGDLDOSCLICVARVPKMRPLPSESEFTTODLOGKITSLDSTMRAMKPGWEDLV 300
QY 301 RRCIOKFHAQHEGESVYAKRHHEVLRQGLAFSQIYRFSLDGTLYAAOTKSKLIRSQI 360
Db 301 RRCIOKFHAQHEGESVYAKRHHEVLRQGLAFSQIYRFSLDGTLYAAOTKSKLIRSQI 360
QY 361 TNEPOLVISLHMLHREONVCVNPDLTGQTMKPLNFISSNSPAHQALCSGNPGQDWTLS 420
Db 361 TNEPOLVISLHMLHREONVCVNPDLTGQTMKPLNFISSNSPAHQALCSGNPGQDWTLS 420
QY 421 SNTNFPINGKEOMGMPMGFRGGSGGNHVSQMOATTPGOSNVALKKNSSQSSPGNPG 480
Db 421 SNTNFPINGKEOMGMPMGFRGGSGGNHVSQMOATTPGOSNVALKKNSSQSSPGNPG 480
QY 481 OPTSMLSPRHRMSPGVAAGSPRIPPOFSFASLSHPVGVSGTGNSSHSTYNSLSMALQAL 540
Db 481 OPTSMLSPRHRMSPGVAAGSPRIPPOFSFASLSHPVGVSGTGNSSHSTYNSLSMALQAL 540
QY 541 SEHGVSLSGSSLASPDLMKGNLONSPVNNMPPPLSKWGSJDSKDCRGLGEPEBEGTGA 600
Db 541 SEHGVSLSGSSLASPDLMKGNLONSPVNNMPPPLSKWGSJDSKDCRGLGEPEBEGTGA 600
QY 601 ESSCHGEQEKETDNPMLPVPVSSERADGOSRLDSKQOTLLDOLLTKSOMPSPLASS 660
Db 601 ESSCHGEQEKETDNPMLPVPVSSERADGOSRLDSKQOTLLDOLLTKSOMPSPLASS 660
QY 661 LSDTNKDSSTGLSPGSGTSGTSLKEKHILHRLDSSSPVDLAKLTAEATGKDLSESS 720
Db 661 LSDTNKDSSTGLSPGSGTSGTSLKEKHILHRLDSSSPVDLAKLTAEATGKDLSESS 720
QY 721 STAPGSEVTIKOEPVSPKKKENALLRYLDKDDTKDIDGLEITPKLERLDSKTDPAASNTK 780
Db 721 STAPGSEVTIKOEPVSPKKKENALLRYLDKDDTKDIDGLEITPKLERLDSKTDPAASNTK 780
QY 781 LIAMKTEKEMSEFPGDOPSELDNLEILLDIONSOPLPDPTRGAGAGVDAKAIT 840
Db 781 LIAMKTEKEMSEFPGDOPSELDNLEILLDIONSOPLPDPTRGAGAGVDAKAIT 840
QY 841 NDLMOJLAENSPTVYGAOCTALRISOSTFNPNRPGOLGRLPNOMLIDITQSPGAG 900
Db 841 NDLMOJLAENSPTVYGAOCTALRISOSTFNPNRPGOLGRLPNOMLIDITQSPGAG 900
QY 901 PPPPIRNSSYVITPOPGMGNOMIGNONLGNSSGTGMITGNSASRPTMPSGEMAPOSSA 960
Db 901 PPPPIRNSSYVITPOPGMGNOMIGNONLGNSSGTGMITGNSASRPTMPSGEMAPOSSA 960
QY 961 VRYTCAATTSAMNRPVGGGIRNPASIPMRPSOPOROTLOSQOVNITGSPSELEMNGG 1020
Db 961 VRYTCAATTSAMNRPVGGGIRNPASIPMRPSOPOROTLOSQOVNITGSPSELEMNGG 1020
QY 1021 POYSQOQAPPNQOTAPWPESTILPIDQASFASQNRQFEGSSPDLLCPHRAESPSDEGALL 1080
Db 1021 POYSQOQAPPNQOTAPWPESTILPIDQASFASQNRQFEGSSPDLLCPHRAESPSDEGALL 1080
QY 1081 DQLYLALRNFDELEIDRALGIPELVSOQAVDPEQFSODSNIMLEQKAPVFPQOYASQ 1140

```

```

Db 1081 DQLYLALRNFDELEIDRALGIPELVSOQAVDPEQFSODSNIMLEQKAPVFPQOYASQ 1140
QY 1141 AQMAOGSSYSPMODPNFHTMGQRPSTATLRMQOPRGLRPTGLVONOPNOLRLQOHLQAO 1200
Db 1141 AQMAOGSSYSPMODPNFHTMGQRPSTATLRMQOPRGLRPTGLVONOPNOLRLQOHLQAO 1200
QY 1201 QNRQPLMNOISVSNVNLTRPGVPTQAPINAOMLAQOREIILNHLRQORHMOQOOVOO 1260
Db 1201 QNRQPLMNOISVSNVNLTRPGVPTQAPINAOMLAQOREIILNHLRQORHMOQOOVOO 1260
QY 1261 RTLMRGQGLNMTPSVAVPSGMPATMSNPRIPQANNOQFPFPNXYGISOQPDGFGTATT 1320
Db 1261 RTLMRGQGLNMTPSVAVPSGMPATMSNPRIPQANNOQFPFPNXYGISOQPDGFGTATT 1320
QY 1321 POSPLMSPRHAHTQSPMDOOQANPAYQAPSDINGMAOGMGNMSFSDQSPPHFQOQAN 1380
Db 1321 POSPLMSPRHAHTQSPMDOOQANPAYQAPSDINGMAOGMGNMSFSDQSPPHFQOQAN 1380
QY 1381 TSMYSNNMNINVSMAATNTGSSMSMNOQTGOISMTSVTSVTSGLSWMGPQVNDPALRG 1440
Db 1381 TSMYSNNMNINVSMAATNTGSSMSMNOQTGOISMTSVTSVTSGLSWMGPQVNDPALRG 1440
QY 1441 NLEPNOLPGMDATKQEGDTRKTC 1464
Db 1441 NLEPNOLPGMDATKQEGDTRKTC 1464

RESULT 2
US-08-891-640-3
; Sequence 3, Application US/08891640
; Patent No. 6268173
; GENERAL INFORMATION:
; APPLICANT: Chambon, Pierre
; APPLICANT: Gronemeyer, Hinrich
; APPLICANT: Voegel, Johannes
; TITLE OF INVENTION: Transcriptional Intermediary Factor-2
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, NW, Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/891,640
; FILING DATE: Herewith
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 60/021,247
; FILING DATE: 12-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Steffe, Eric K.
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1383.0130001/EKS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1036 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: not relevant
; MOLECULE TYPE: peptide
; US-08-891-640-3

```

Query Match 13.6%; Score 1037.5; DB 4; Length 1036;  
 Best Local Similarity 28.7%; Pred. No. 2,3e-66;  
 Matches 346; Conservative 179; Mismatches 339; Indels 343; Gaps 52;

```

OY 395 LNPISNSPQAOLCSGN--PGQDWTLSNINFPINGKREO-----MGMPMGPF 441
DB 10 VNP--SISPAIGVARSSTLPPSNMNVSTRIN-----RQSSDHLSSHSSNSQSGF 61
OY 442 GSGGGMNVSGM-----OATTPQGSNYALKNPSSQSSPGMNPQPTSMLEPRHMSDGV 496
DB 62 GCSRQSLVAVNALKKGAASS--QSSKPSLNLNPPMEGIGISLAQ---FMSPRQVTSGL 117
OY 497 AGSPRIIPSOQSP--AGSLHSPYGVCS--TGNSHSYTSSNLALALSEGHVSLGSSLA 553
DB 118 ATRPMPNNSFPNISTLSLSPVGMTSSACNNNNRYSNIPVTSLOGMDEGPNNSVGFSA 177
OY 554 SPDLMGNLQNSPVNMPNPLSKMGLSDKCFGLYGEPEEGTQAESSCHPGQKET- 612
DB 178 SPVLROMSSQNSPSRLNTP--AKAESKDKKEIASLNLNEMI-----QSDNSSDGRPLDSG 231
OY 613 ---NDPNLPPAVSSERADQGRSLHDSKQTKLLQLLTTKSD--QMBPSPLASSLD---- 663
DB 232 LLHNNDRLL-----SDGSKY--SQTSHKLVLQTLTTAEQQLRHADIDTSCKDVLSCT 281
OY 664 ---TKDSTGSLPGSGSTGTSLKEKHKILHLRLDSSSPVDLAKTAEATGKDLSSSS 721
DB 282 GTSNSASANSSSGSCPSHSSSLTARHKILHLRLQ--GSPDITLTSVPEKKD--SASTSV 339
OY 722 TAPG---SEVTKOEPVPPKKE---NALRYLIDKDD---TKDIGLPEITPLERLD 770
DB 340 SVTGOVGNSSIKLELDASKKESKDHOLRLDKDEKDLSTYNLISLDYKVAVERKE 399
OY 771 SKTDPASNTKLIAMKTEKEMSEFEGDQSGSELDNLEELIDQLN--SQLPOLPPTRG 829
DB 400 -QMPBCNTPPTPKATPEEIKLEAOSQFTADLDQFDQLPTLEKAAQLPGICETDRMDG 458
OY 830 PAGSYDKAIIINDLMQLTAENSPYVGAQKTALRISQSTFNPRGQLRLLPQNLPL 889
DB 459 AVTSV-----TIKS 467
OY 890 DITLQSPGAGFPPIRNSPSYVLPORGMMGNQGNIGNSGTMIGNSASRPTM 949
DB 468 EITIKS-----ELIP-----ASLQSAIARP-- 487
OY 950 PSGEAPQSSAVRYTCAATTAMNRPVCGMIRNPAISIPMRSSQPGQRTLQSOVANI 1009
DB 488 -----TSRLNR-----LP----- 495
OY 1010 GPSELEMMAGROYSOQAPRNOTAPRP--ESTLPIDQASFAQNRQPPGSSPDLLCPHP 1068
DB 496 ---ELELEAIDNQGQPGPT--GDQIPMTNNTVTAIINQ---SKSEDCISSQDELCPPT 547
OY 1069 AASPDEGALIDLP--YIALRNFDGLEIDRALGIPELVISOQAVD--PROFSSQDSN- 1123
DB 548 TVEGRNDEKALIBOLVSLSGDELELADRALGIDIKLV--QGGGLDVLSSEFPQOATP 606
OY 1124 -TMEOKARVEPQOYASQAOQAQGSYSMPDQNFHTMGORPSYATLRMQPRP----- 1174
DB 607 PLMEERENLVSQPYSSPPTAN-----LPSPQGMVROKPSLGMVYVTPPRGAFSPG 661
OY 1175 -GLRPTGLVON---QPNLRLOLQRLQAO-----QNRQPLMNOISNVSNULTRPFG- 1223
DB 662 MGMPROTLPNPPAPRNQLRLQDQRLQGOOQLHQNROAILNPPAARPAVGINRRSQ 721
OY 1224 --VTQAPINQMLAORREILNOLHROHMOQOQVOOQRTLMRGQ--GLMUTSKYAP 1279
DB 722 QOITPQRLNOMLAORREILXSOHROROLIQ---QORALIMQOQSGNULPSP- 773
OY 1280 SGMPATKSNPRIPOANAOQPPRPYGISQOPDPEFTG---ATTPQSPF----- 1325
DB 774 SGLPVQOTGNPRLPQAGPOQFPYRPYGTNPGRPASTSPFQSLAANPEASLANRRSNVSR 833

```

```

OY 1326 -----MSPRAHNT-----QSPM 1337
DB 834 GMTGINGGQFGTGIMPQMOQNVFOYPGAGMPVQGEANFAPSLSPSSSNWMDIPPOSL 893
OY 1338 MOOSQANPAYQAPSDINGNAQMGNSMFSQ--QSPHFGQOATMYSNNNINVSMA 1395
DB 894 LOOTPPASQOSP--DMKAMQGAIGNNVFSQAVQNP---TPAQPYV--NNMSTVSMA 948
OY 1396 TMTGMSNNQMTGOISMTSVTSSTGSLSMGPQVNDPALRGNLPPNOLPGMDMIQ 1455
DB 949 GGNTVQNMNPPMAQOMQ---SLQMPGMVTCPEQINDPALRHGILCNQSLTDLKT 1005
OY 1456 EGDTRK 1462
DB 1006 EADGTQ 1012

```

```

RESULT 3
US-08-885-291-55
; Sequence 55, Application US/08885291A
; Patent No. 6057125
; GENERAL INFORMATION:
; APPLICANT: Takahashi, Joseph S.
; APPLICANT: Turek, Fred W.
; APPLICANT: Pinto, Lawrence H.
; TITLE OF INVENTION: CLOCK GENE AND GENE PRODUCT
; FILE REFERENCE: 0290-5
; CURRENT APPLICATION NUMBER: US/08/885,291A
; EARLIER FILING DATE: 1997-06-30
; EARLIER APPLICATION NUMBER: 08/816,693
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 55
; LENGTH: 846
; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-885-291-55

```

Query Match 4.2%; Score 317; DB 3; Length 846;  
 Best Local Similarity 21.5%; Pred. No. 2e-14;  
 Matches 176; Conservative 129; Mismatches 316; Indels 196; Gaps 39;

```

OY 31 KNTTEKRNDEQNKYTELELIFANFNDIDNFKRDPKCAILKETWQIROIKQEKAA 90
DB 39 RUKSEKRRDQFNVLKELGSMPCNAR-----KMDKSVLQKSIDFLR--KKHEITA 89
OY 91 ANIDVOKSDVSSGQGVIDKDALGPMLELDGFFVNVNLEGVNVFVSENVQYLKRN 150
DB 90 OSDASEI--RQDKKPT--FLSNEEPTQLMLELDGFFLAIMDGSIIYVESVTSLEHL 145
OY 151 QELMKNKSVSYILHGDHTEFKNLKPSIYVSGSMGSGPPR--RNSHFPNCML----- 203
DB 146 PSDLDVQSGIFNTPGSEHSYVK--ILSTHLESDBLTPRYLXSKNQLECCIMLGITDP 204
OY 204 -----VKPLPDSSEEGHDN---QEAHQ--KYETMOQFA---VSQPKSTK 239
DB 205 KEPSTYEVYKFIKGNKSLNSVSSAHNGFEQTIQRTHRPSYEDRVCFATVRLATPQFIK 264
OY 240 EEGEDLOSLCLVARVRPKERPVLPSSSEFTTRDQDQKITSLDSTYRAMKRGWEDL 299
DB 265 E-----MCTV-----EEP-----NEEFTSRSLKWKFLFD--HRAPPIIGVLPF 302
OY 300 VRCIOKFAHQEGESVSYAKRHHEVLROGLAFQIYRFSLSDGTLVAQTKSKLIRSQ 359
DB 303 EVLGTSGDYIYVDDLENAKCHEN--LMQYKKGSCYIRFLTKGQOWIMQIHHYITVHQ 361
OY 360 TTNEPOLVLSLMLHREONVCYMPNDLTGQITGKPLNPISNSPAHQALCSGNPGQDML 419
DB 362 WNSRPEFIVCTHTVVSVAEVAERREELG-----IEESLPTAADKSDQSSDNRNI 412

```

```

QY 420 SSINFPINGPKQMGMPKRGFGSGMNHVSGMATTPOGSHYALKMNS-PSQSSPGMN 478
DB 413 NT-----VSLKEALERFDHS-----PTPSASSRSSKSSHTAVSPSSST 451
QY 479 PCQ-PTSMLSPRHRMSPGVAGSPRIPPSQFSPAGSLHSPVGVSSYGNSSYNSLSNAL 537
DB 452 PKIPIPTDTS-----PPRQHLPA---HEKM-----VORSSSFSSQISNS- 487
QY 538 QALSEGHVSLGSSSLASPDLMKGNLONSPVNMNPPPLS-----KMGSLDS-KDCFGLYG 590
DB 488 -----QSVGSSLTQP---VMSQATNLPT---PQMGSGQFSAQIGAMQHLLKD----- 528
QY 591 EPSEGTGGAESSCHPEQKETNDPMLPPAVSSERADG-OSRLHDSK-----GQTKLLDL 644
DB 529 -QLEQRTMIEANIH-RQOEELR--KIQEQLQWVGQGLQMFLOQSNPGLNFGSVOLSSG 584
QY 645 LTTKSDQMEPSPLASSLSPTNKDSTGSLPGS--GSTHGTSLKEKHKLILHRLLDSSSPVDL 703
DB 585 NSSNIQOLAPINNQGVPVPTNQIOSGMNTHGIGTT-----QHMTQQOQTLOSTSTQSOQ 637
QY 704 AKLTAEATGKDLQSESSS--TAPGSEVTTIKQEPVSPKKENALLRYLLDKDRTKDIGLPE 761
DB 638 NVLSGHSQOTSLPSQOTSTIATPLANTMVISQPAAGSM-----VQIPS 680
QY 762 ITPKLERLDSKTPASNTKLIAMKTEKESKSEPPGDO 798
DB 681 SMPQ-----NSTQSAAVTTF---TODRQIRFSGQOO 708

```

RESULT 4  
US-09-107-847-2

Sequence 2, Application US/09107847

Patent No. 6100062

GENERAL INFORMATION:

APPLICANT: DUCKWORTH, DAVID

APPLICANT: MICHALOVICH, DAVID

TITLE OF INVENTION: NOVEL USE

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:

ADDRESSEE: Ratner & Prestia

STREET: P.O. Box 980

CITY: Valley Forge

STATE: PA

COUNTRY: USA

ZIP: 19482

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/107, 847

FILING DATE: 30-JUN-1998

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: EP 97304996.8

FILING DATE: 08-JUL-1997

ATTORNEY/AGENT INFORMATION:

NAME: Prestia, Paul F.

REGISTRATION NUMBER: 23,031

REFERENCE/DOCKET NUMBER: GH-30003

TELECOMMUNICATION INFORMATION:

TELEPHONE: 610-407-0700

TELEFAX: 610-407-0701

TELEX:

INFORMATION FOR SEQ. ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 846 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-107-847-2

Query Match 4.2%; Score 317; DB 3; Length 846;  
Best Local Similarity 21.5%; Pred. No. 26-14;  
Matches 176; Conservative 129; Mismatches 316; Indels 196; Gaps 39;

```

QY 31 KRTEKRNROENKYTELELAELIFANFNIDNENFPDKCALIKETVKQIRQIKDEKKA 90
DB 39 RNSSEKRRQDFVNLRIKELGSMPLGNAR-----KMDKSTVLDKSIDFLR--KHKEIYA 89
QY 91 AANIDEVQKSDVSTGQGVIDKDALPMLLEALDGFEEFVNVNEGNAVVFSENYQVLRN 150
DB 90 QSDASEI-RQDWKPT---FLSNEEPTQLMLEALDGFFLAMTQDSTIYSESTSLLEHL 145
QY 151 QEELMKSVYSILHVDHTEFVKNLLPKSTVNGSGSGEPPR-RNSHTFNCRLM----- 203
DB 146 PSDLVQOSTIFNFTPEGEHEVYK-ILSTHLLBESDILTPEYIAKKNODEFCCHMLRTIDP 204
QY 204 -----VKRLPDSEEGHDN-----QEAHQ-KYETMGFA---VSQPKSIK 239
DB 205 KEPSTVEYVYKFIQNFKSLNSVSSAHNGFEGTIOFTHRPSYEDRVCFVATVRLATPQFIK 264
QY 240 EBEDEDIQCCLICVARRVPKREPVLPSSSEFTTRQDLOGKITSLDSTYRAAMKPGMEDL 299
DB 265 E-----MCTV-----EEP-----NEEFTSRHSLKMFLELD--HRAPPIGYLDP 302
QY 300 VRRCIOKFNAHGESESVYAKRHNHVELRQGLAFSOLYRFSLSDGTLLVAQTKSKLIRSQ 359
DB 303 EVLGTSQDYDYHVDLLENLAKCHEH-LMQYKKKSCSYRFLRFGQGMWMLQFTYITTHQ 361
QY 360 TTNEPOLVITSLMLHREQVNCVNPDLTGOTMCKPLNPJSSNSPAHQALCSGPNQDML 419
DB 362 WNSRPEFIVCTHTVVSYAERARREELG-----IEESLPETAAKSDQSDGSDNRI 412
QY 420 SSINFPINGPKQMGMPKRGFGSGMNHVSGMATTPOGSHYALKMNS-PSQSSPGMN 478
DB 413 NT-----VSLKEALERFDHS-----PTPSASSRSSKSSHTAVSPSSST 451
QY 479 PCQ-PTSMLSPRHRMSPGVAGSPRIPPSQFSPAGSLHSPVGVSSYGNSSYNSLSNAL 537
DB 452 PKIPIPTDTS-----PPRQHLPA---HEKM-----VORSSSFSSQISNS- 487
QY 538 QALSEGHVSLGSSSLASPDLMKGNLONSPVNMNPPPLS-----KMGSLDS-KDCFGLYG 590
DB 488 -----QSVGSSLTQP---VMSQATNLPT---PQMGSGQFSAQIGAMQHLLKD----- 528
QY 591 EPSEGTGGAESSCHPEQKETNDPMLPPAVSSERADG-OSRLHDSK-----GQTKLLDL 644
DB 529 -QLEQRTMIEANIH-RQOEELR--KIQEQLQWVGQGLQMFLOQSNPGLNFGSVOLSSG 584
QY 645 LTTKSDQMEPSPLASSLSPTNKDSTGSLPGS--GSTHGTSLKEKHKLILHRLLDSSSPVDL 703
DB 585 NSSNIQOLAPINNQGVPVPTNQIOSGMNTHGIGTT-----QHMTQQOQTLOSTSTQSOQ 637
QY 704 AKLTAEATGKDLQSESSS--TAPGSEVTTIKQEPVSPKKENALLRYLLDKDRTKDIGLPE 761
DB 638 NVLSGHSQOTSLPSQOTSTIATPLANTMVISQPAAGSM-----VQIPS 680
QY 762 ITPKLERLDSKTPASNTKLIAMKTEKESKSEPPGDO 798
DB 681 SMPQ-----NSTQSAAVTTF---TODRQIRFSGQOO 708

```

RESULT 5  
US-09-496-672-55

Sequence 55, Application US/09496672

Patent No. 6291429

GENERAL INFORMATION:

APPLICANT: Takahashi, Joseph S.

APPLICANT: Turek, Fred W.

APPLICANT: Pinto, Lawrence H.

TITLE OF INVENTION: CLOCK GENE AND GENE PRODUCT

FILE REFERENCE: 0290-5



CURRENT APPLICATION NUMBER: US/09/496,672  
CURRENT FILING DATE: 2000-02-03  
PRIOR APPLICATION NUMBER: 08/885,291  
PRIOR FILING DATE: 1997-06-30  
PRIOR APPLICATION NUMBER: 08/816,693  
PRIOR FILING DATE: 1997-03-13  
NUMBER OF SEQ ID NOS: 55  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 55  
LENGTH: 846  
TYPE: PRF  
ORGANISM: Homo sapiens  
US-09-496-672-55

Query Match 4.2% Score 317; DB 4; Length 846;  
Best Local Similarity 21.5%; Pred. No. 2e-14;  
Matches 176; Conservative 129; Mismatches 316; Indels 196; Gaps 39;

QY 31 KRTEKRRREOENKYEIELLIFANFNDIDNFKPKDCAILKETVQIROIKEOEKAA 90  
DB 39 RKNSEKRRRODPVNLIKELGSMLEGNAR-----KMDKSTVLOKSIDFLR--KHKEITTA 89  
QY 91 AANIDEVOKSDVSTGGGVIDKDALGPMLEALDGFVVNLGNNVVFSENVTOYLRYN 150  
DB 90 QSDASEI-RODMKPT---FLSNEEFTQMLEALDGFALIMTDSIIYVSEVSTLLEHL 145  
QY 151 QEELMNKSVSILHVGDTHEVKNILPKRSYVNGSWSGEPPR-RNSHTFNCRL----- 203  
DB 146 PSDLDVDSIFNFIEGEHSEYK--ILSTHILESLSLTPPEYLKSNQLEFFCCHMLRGITDP 204  
QY 204 -----VKPLPDEEGHND-----QEAHQ-KYETWQCFP-----VSOPKSIK 239  
DB 205 KEPSTYEVKFIKFNKLSNVSSAHNCFEETIQRTHRPSTEDVNCFAIVRLATPOKIK 264  
QY 240 EEGEDLOSLICVARRVPMKERPVLPSSSEFTTQDLOGKITSIDTSTMARAMPGWEDL 299  
DB 265 E-----MCTV-----EEP-----NEFTSRHSLEMKFLFD---HRAPPIIGYLPF 302  
QY 300 VRRCIOKFNHOGHEESVYAKRHHHEVLRQGLAFSQIYRESLSGTLVAOTKSKLRSQ 359  
DB 303 EVLGTSGDYVHVDLEMLACHEH-LMÖYKGKSCYYRFLTKGOQWIMLTGTHYYITYVHQ 361  
QY 360 TTNEPOLVLSLHMLHREQNVCVMPDLTGOTMGKPLNPISNSPAHQALCSGNPGQDWTL 419  
DB 362 WNSRPERIVCHYVVAEVAERRELG-----IESLLETADKQSDGSDMKRI 412  
QY 420 SSNINPDPKQMGPMGRFGSGGMNHYSGMATTPOGSNYALKNNS-PSOSSPGMN 478  
DB 413 NT-----VSLKEALERFDHS-----PTPSASSRSRSKSSHTAVSDPST 451  
QY 479 PGQ-PTSMLSRHRMSGCVAGSPRIIPSOFSPPASLSHPVCGSTGNSHTYSSLNAL 537  
DB 452 PTKIPPTJST-----PRQHLPA---HEKM-----VORRSFSSOSINS- 487  
QY 538 QALSEGHVGLSSGLASPDLMGNLONSPVMPNPPLS-----KMGSLDS-KOCFGLYG 590  
DB 488 -----QSVGSSLTP---VMSQATNLP---PQMSQOFQFSQDLGAMQHLKD----- 528  
QY 591 EPSEGTGAESSCHPEOKETNDPNLPPAVSSERADG-QSRLHDSK-----GOTKLLOL 644  
DB 529 -QLEQRTIRMEANIH-RQOEELR--KIOEQLOMWHGGLQMFLOOSNGMLFGSVQSSG 584  
QY 645 LTTKSDOMESPPLASSLSDTNKSDGSLPGS-GSTHGTSLKEKHKLHRLLODSSSPVDL 703  
DB 585 NSSNIQOLAPINMOGVVPTNIOQSGMNTGHTGT-----QHMIOQOOLSTOSTSOQ 637  
QY 704 AKLFAEATGKDLQESS--TAPGEVTLKOEVPSPKKKENALLRYLLDKDXTDIDIGPE 761  
DB 638 NVLSGHQOQSTSLPSQOSTLAPLNTWVLSQAPAGSK-----VOIIPS 680  
QY 762 ITPKLERIDSKTDPASNTKLIAKTEKEEMSFEEGDQ 798  
DB 762 ITPKLERIDSKTDPASNTKLIAKTEKEEMSFEEGDQ 798

DB 681 SMPQ-----NSTOSAAVTTF-----TQDROIRESGQOQ 708

## RESULT 6

US-08-816-693A-2  
Sequence 2, Application US/08816693A  
Patent No. 5874241  
GENERAL INFORMATION:  
APPLICANT: Takahashi, Joseph S  
APPLICANT: Turek, Fred W  
APPLICANT: Pinto, Lawrence H  
TITLE OF INVENTION: Clock Gene and Gene Product  
NUMBER OF SEQUENCES: 53  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dressler, Rocky, Milanow & Katz  
STREET: Two Prudential Plaza, Suite 4700  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/816,693A  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 5874241thrup, Thomas E  
REGISTRATION NUMBER: 33,268  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-616-5400  
TELEFAX: 312-616-5460  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 855 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-816-693A-2

Query Match 4.1% Score 309.5; DB 2; Length 855;  
Best Local Similarity 21.7%; Pred. No. 7e-14;  
Matches 150; Conservative 107; Mismatches 254; Indels 181; Gaps 27;

QY 31 KRTEKRRREOENKYEIELLIFANFNDIDNFKPKDCAILKETVQIROIKEOEKAA 90  
DB 39 RKNSEKRRRODPVNLIKELGSMLEGNAR-----KMDKSTVLOKSIDFLR--KHKEITTA 89  
QY 91 AANIDEVOKSDVSTGGGVIDKDALGPMLEALDGFVVNLGNNVVFSENVTOYLRYN 150  
DB 90 QSDASEI-RODMKPT---FLSNEEFTQMLEALDGFALIMTDSIIYVSEVSTLLEHL 145  
QY 151 QEELMNKSVSILHVGDTHEVKNILPKRSYVNGSWSGEPPR-RNSHTFNCRL----- 203  
DB 146 PSDLDVDSIFNFIEGEHSEYK--ILSTHILESLSLTPPEYLKSNQLEFFCCHMLRGITDP 204  
QY 204 -----VKPLPDEEGHND-----QEAHQ-KYETWQCFP-----VSOPKSIK 239  
DB 205 KEPSTYEVKFIKFNKLSNVSTHNGFEETIQRTHRPSTEDVNCFAIVRLATPOKIK 264  
QY 240 EEGEDLOSLICVARRVPMKERPVLPSSSEFTTQDLOGKITSIDTSTMARAMPGWEDL 299  
DB 265 E-----MCTV-----EEP-----NEFTSRHSLEMKFLFD---HRAPPIIGYLPF 302  
QY 300 VRRCIOKFNHOGHEESVYAKRHHHEVLRQGLAFSQIYRESLSGTLVAOTKSKLRSQ 359  
DB 303 EVLGTSGDYVHVDLEMLACHEH-LMÖYKGKSCYYRFLTKGOQWIMLTGTHYYITYVHQ 361  
QY 360 TTNEPOLVLSLHMLHREQNVCVMPDLTGOTMGKPLNPISNSPAHQALCSGNPGQDWTL 419

```

Db      362  WNSRPEFLVICHHTYVSYAEVAERRELG-----IEESLPETADKSDQSDSNRI 412
Oy      420  SSNINPPINGPEKDMGMEGRFGSGGNNHVSQMGAATPPGGSNYALKMNS-PSQSSFGMN 478
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db      413  NT-----VSLKEALERFDHS-----PTPSASRSRSKRSHTAVIDPSSST 451
Oy      419  PGQ-PTSMLSLSP-----RRHMSGVAGSRRIPPSQSPSPAGSLSHSPVGYCST 523
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      452  PTKIETDITSTPEPROHLPAHEKMTGRSSFSQSSINSQSVGSLTPAPMSQANLPI--PQ 509
Oy      524  GNSHSYTNSSNALQALASE-----GHGVSLGSSL 552
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      510  GMSQGFQFAQLGAMQHLKDQLLEORTMTLEAHIRHQEELKIQEQDLMVHGQSLQMTLQ 569
Oy      553  ASPDLKMGNLONS-----PVNNMNP--PLSKMGSLDSKDCFLGYEPESEG----- 595
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db      570  SMLGINFQSYQLSSGNSNIQLUTLPVNMQGVVPANVOVS-----GHISGQIMIQ 619
Oy      596  -----TTGAESSCHPGEDEKETINDPLPA 620
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      620  QQTLLSTSTQSQSQSVMSGHSGQSQSLSPQSPQS 651

```

```

RESULT 7
US-08-885-291-2
; Sequence 2, Application US/08885291A
; Patent No. 6057125
; GENERAL INFORMATION:
; APPLICANT: Takahashi, Joseph S.
; APPLICANT: Turek, Fred W.
; APPLICANT: Pinto, Lawrence H.
; TITLE OF INVENTION: CLOCK GENE AND GENE PRODUCT
; FILE REFERENCE: 0290-5
; CURRENT APPLICATION NUMBER: US/08/885,291A
; CURRENT FILING DATE: 1997-06-30
; EARLIER APPLICATION NUMBER: 08/816,693
; EARLIER FILING DATE: 1997-03-13
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Mus musculus
US-08-885-291-2

```

```

Query Match Similarity      4.1%; Score 309.5; DB 3; Length 855;
Best Local Similarity       21.7%; Pred. No. 7e-14;
Matches 150; Conservative 107; Mismatches 254; Indels 181; Gaps 27.;

QY      31 KRTEKRNREOENKYEEETLAELIFANFNIDNMFNEPKDCAIKETVKROIKEDEKA 90
        : |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db       39 RKKSEKKRRDDQNVNLIKEIGSMIPGNAR-----KMDKSTYLQKSIDLR--KHKETTA 89

QY      91 AANIDEVOKSDVSSTGOGVIDKDALGPMLLEALDGEFFVNLEGNNVFSEVNTOTLYRN 150
        : |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db       90 QSDASTI-RQDKMPT---PLSNFEFTQLMEALDGEFLAIMDGSIITYSESVTSLIELH 145

QY      151 OELMNKSYSVLHWGDHTEFYKNLLPKSIVNGSGNSGEPFR-RNSHTENCRL----- 203
        : |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db       146 PSDLVDQSIFNFIRECEHSYVK-IISTHLISDSITPEYLSKNOJLFCCHMLRGTIDP 204

QY      204 -----VKRPDPSEBECHDN-----QEAHQ-KYTEMOCEA-----VSOPKSTK 239
        : |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db       205 KEPSTYEYREFIGNFSLSVSTSHNGCEGTIOIRHNRESYEDRVCFATVRLATPOQFK 264

QY      240 EREGEDLOSLICVARVPKPERVLPRSSSFTRRDLOGKITSLDTSTYRAAMKPWEML 299
        : |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db       265 E-----MCTV-----EER-----NEEFSTRHSILEMKFFLLD---HRAPPIIGLPF 302

QY      300 VNRCLCQFPNAOHGESVSAYAKRHNHVEULROGLAFSQIYRFSLSDGLVAQOKSKILRQ 359
        : |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

```

Db      303  EVLGSIGSDYHHVADDELYLACHHEH-LM0YKKGKSCYYRFLTKQGMIMVQTHYYIITHQ 361
Qy      360  TTNEPOLVLSLMLHREONVCVMNBDLTGOTMGRPLNPISNSPFAHOLCSGNPGDMTL 419
Db      362  WNSRPEFIVCTHTVVSYLEVAERREILG-----IEESLPETAADKSDGSDPNRI 412
Qy      420  SSNINPFLNGKEQMGMGMGFFGSGGNNHVSQMATTPOGSNVALKMS-PSGSSGMN 478
Db      413  NT-----VSLKALEFFDHS-----PPSASSSSSKSSSHTAVDSSST 451
Qy      479  PGO-TTSMLSR-----RHMSPGVAGSPRIPSPGSLHSPVGYCST 523
Db      452  PKPIPTDTSPPROHLRPHKMTQRRSSFSSQINSQSVGSLTPRPMASQANLP-PI 509
Qy      524  GNSHSYTNSSNALQALSE-----GHEVSLGSSSL 552
Db      510  GMSQGFQFAGQGMQHLKDLQEQRTMYEANIHRQOEELKIQEQQLVMHGQGLQMP 568
Qy      553  ASPDLKMGNLONS-----PYNNMP--PLSKMSGLDSDKCFGLYGPSPG 595
Db      570  SNPGLNFGSVDLSGNSNIQQLTPVNMGGQVVPANOVQS-----GHISGQHMIO 619
Qy      596  -----TTGOAESCHPGEQKETNDPLPLPA 620
Db      620  QQTLOSTSTQDSQGSVMGSHSQQTSLSPQTS 651

```

```

RESULT      8
US-09-496-672-2
; Sequence 2, Application US/09496672
; Patent No. 6291429
; GENERAL INFORMATION:
; APPLICANT: Takahashi, Joseph S.
; APPLICANT: Turek, Fred W.
; APPLICANT: Pinto, Lawrence H.
; TITLE OF INVENTION: CLOCK GENE AND GENE PRODUCT
; FILE REFERENCE: 0290-5
; CURRENT APPLICATION NUMBER: US/09/496,672
; CURRENT FILING DATE: 2000-02-03
; PRIOR APPLICATION NUMBER: 08/885,291
; PRIOR FILING DATE: 1997-06-30
; PRIOR APPLICATION NUMBER: 08/816,693
; PRIOR FILING DATE: 1997-03-13
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-496-672-2

```

	Query Match	4.1%	Score 309.5,	DB 4;	Length 855.	
	Best Local Similarity	21.7%	Pred. No. 7e-14;			
	Matches 150;	Conservative 107;	Mismatches 254;	Indels 181;	Gaps 27;	
Oy	31	KRNETKRNNEQGNKKYIEELAEILFANFNNDIDNFMNRPDKCALIKETVQIROIKDOEKAA	90			
	:	: : : : :	: : : : :	: : : : :	: : : : :	
Dd	39	RNKSEKRKRDDPNVLILKETIGLSMLPGNAR----	KMDSTYLQSIDLR--KHKETA	89		
Oy	91	AANIDEVOKSDVSSNGGAGVIDDALGPMMLEALDGFPPVNLEGNVAFSEAVTQIARN	150			
	:	: : : : :	: : : : :	: : : : :	: : : : :	
Dd	90	QSDASTI-RQDMKPRT---FLSNBEFPQLMDEALDGFFLAIMDGSITTYSESVTSLBHL	145			
Oy	151	OEILANKSVYSILHVGADTHEFYVKNLLPKRSIVNGSSGGEPR-RNSHRFNCRL-----	203			
	:	: : : : :	: : : : :	: : : : :	: : : : :	
Dd	146	PSDLVDOSIFNFIPREGEHEVYK-LISTHALDESISLTPELYLKSNOLEPCCHMLRGITDP	204			
Oy	204	-----VKPRLDSEEGHDN-----OEAHQ-KYEUMOCFA-----VSOPKRS	239			
	:	: : : : :	: : : : :	: : : : :	: : : : :	
Dd	205	KBPSTVEYRFIGNRKSLTSYSTHNHGEEGIQRNRHSYEDRYCSFVAIVGLAIPROFIK	264			
Oy	240	EKGEDLOGLICVARRRVPKERRPVLESSEFTTRDOLCKITSLOTSTRAAMKRWEMDL	299			

```

Db 265 E-----MCTV-----EEP-----NEEFTSRHSLMKFLFD---HRAPIIGYLPF 302
QY 300 VRRCIOKFNHAEBSYSYAKRHHHEVLRQGLAFSQIYRFSLSDTLVAQTKSLINSQ 359
Db 303 EVLTSGDYHYVDLELACRCHH-LMVGKSKCYRFLTKGQOMWLQTHYITVHQ 361
QY 360 TTNEPOLVLSLMLHREONCVMPNDLTGOTMGKPLNPISNSPAHQALCSGNPGODMTL 419
Db 362 WNSRPEFLVCHTYVSTAYEVAERRELG-----IEESLPETADKSDSDSDMRI 412
QY 420 SSNINFLPGRKQMGWMPGRFGSGGNHVSQMOATTPQGSNVALKMNS-PSQSSPGMN 478
Db 413 NT-----VSLKEALERFDHS-----PTPSASRSRRKSSHTAVDPSSST 451
QY 479 PGQ-PTMSLP-----RRMSPCVAGSPRIIPSPQSPASSHSPPVCVCSST 523
Db 452 PTKLPDTSTPPROHLPRAHEKMTORSSFSQSINSQSVGSPSLQNPAMSQAANLPI--PQ 509
QY 524 GNSHYTSSSLNALQALSE-----GHSVSLGSSL 552
Db 510 GMSQFQSAQGLAMQHLKQDLEQRTMTEANIHRQOELRKIQEOLQVHGOGLQMLQ 569
QY 553 ASDPLKMGNIQNS-----PVNMP--PLSKMGLSDKDCFGLYGEPSESG----- 595
Db 570 SNPLNGSVGLSSGNSNIQQLTPVNMGGQVVPANQVQS-----GHISTGQHMIQ 619
QY 596 -----TTGQAESCHPEQEKETNDPULPPA 620
Db 620 QOTLQSTSTQSOOSQSVMSGSHSQOTSLPQSTPS 651

```

# RESULT 9

```

US-08-785-241-4
: Sequence 4, Application US/08785241
: Patent No. 5695963
: GENERAL INFORMATION:
: APPLICANT: McKnight, Steven L.
: APPLICANT: Tian, Hui
: TITLE OF INVENTION: Endothelial PAS Domain Protein
: NUMBER OF SEQUENCES: 7
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
: STREET: 268 BUSH STREET, SUITE 3200
: CITY: SAN FRANCISCO
: STATE: CALIFORNIA
: COUNTRY: USA
: ZIP: 94104
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/785,241
: FILING DATE: 17-JAN-1997
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: OSMAN, RICHARD A
: REGISTRATION NUMBER: 36,627
: REFERENCE/DOCKET NUMBER: UTSD:1229
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 343-4341
: TELEFAX: (415) 343-4342
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 870 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: US-08-785-241-4

```

```

Query Match 3.8%; Score 292; DB 1; Length 870;
Best Local Similarity 20.6%; Pred. No. 1.3e-12;
Matches 217; Conservative 138; Mismatches 406; Indels 294; Gaps 46;

QY 10 DPSRAETRRKRECPDGLSPKRNTEKRRNREONVYIEEL-ELIFANFNIDNENFKPD 68
Db 6 EKKRSSERRKE-----KSDAARCRSKSETEVEFLAHELPLP-----SVSSHLD 52
QY 69 KCAILKETVYKQIQTKEQKAAANIDEYKSDVSTGGGVIDKALCPMLBALDGEFE 128
Db 53 KASIMRLEISFLRTHKLLSYVSENESEAEAD-----QOMNLVYTKALEGFA 100
QY 129 VVNLGNVVFVSENYTOYLRYNOEELMKSVSYLIHVDGHEFVNKLPSKLVNCGSMG 188
Db 101 VVTQGDMLFLESENISKPEGLTQVELTGHSTIDFTGPCHEIRENL--SKNKGSG- 156
QY 189 EPPRRNSHTFNCRLMLVKPLPDSEEGHNDQEAHQYETMQCPAVSQPSIKEEG----- 242
Db 157 -----KSKDMSTREDFFMKMKCTVTRNGRTVNLKSAITWYL 193
QY 243 -----EDLQSLICVARRVPMKERVLP--SSESFTTRQDLOK 279
Db 194 HCTGQVYVYVNCPRPHNSLQGYKEPLSLCLIMCEPIQHPSHMDIPLSKTYFLSRSHMDK 253
QY 280 ITSLSPTSTKRAAMKPGW--EDLVRCIQKFNQHGSESVYAKRHHHEVLRQGLAFSQIY 337
Db 254 FYVCD--RITELIGYHEEELGRSAVEFYALDSENM--KSHQNLCTGQVQVSGQY 307
QY 338 RFLSDGTLVAAQTKSKILRSQTNQTNPOLVLSLH--MLHREONCVMPNDLTGOTMGRP- 394
Db 308 RMLAKHGQVWLEETQGYIYNPRNLQPOCIMQVNVVLSIEENDVYFSMDQI--ESLFRH 366
QY 395 --LNPISNSPAHQALCSGNPGQDMLSSNINFLPGRKE-----QMGMDG----- 439
Db 367 LMANNSIFDSS-----GKG-AVSEKSNFLTKKEPEELAQLPAPGDAITSL 414
QY 440 RFGSGGGMNHVSQMOATTPQGSNVYALKMNSP--QSSPGMPCQPTSMLSPRHRMSPVAG 498
Db 415 DFGNMFESSAYGKAIIIPPSQPMATELRSHTQSEASLPFTV-----PQAAA 464
QY 499 SPRIPQSPAGSLSPVGVCSSTGNSHVSNSLNLQALSEGHVSLGSSLASPLK 558
Db 465 POSTTPSATSSSS-----CSTPNSEPDYITSLDNLK-----LEVYIKLAMQTE 510
QY 559 MGNLNSPVNMPNPLSKMGLSDKDCFGLYGEPSEGTGQAESCHPEQEKETNDPULP 618
Db 511 AKDCSTQTDFFNELDLETLAPY-----IPMDGEDFQLSPIC--PEERLAEINPQST 559
QY 619 PAVSSERADGQSRHDSKQGT-----KILQLITTSQDNEPPLASS 660
Db 560 PQ-----HCFSAMTNIQPLAPVAPRHSFLLDKFOQOLESKTEPEHRPMSSI 607
QY 661 LSDTKDSTGSLP--GSGSTHGTSLKEKKKLIHLRLDSSSPVYDLAKTLATATGKDSQ 717
Db 608 FFDAG--SKASLPPCCGQASTPLSSMGGRN-----TQMPRPPLHFGPTKAAVG--DQRT 659
QY 718 ESSSTAPGSEVYIKQEPVSPKKRENALLRYLLDKDQTDIGL--PEI--TPKLERLDSKTD 774
Db 660 EFLGAAP-----LGPVSPPHVSTFKTR-----SAGFGARGDVLSPAVVAL----- 702
QY 775 PASNFKLIAMKTEKEMSEF--GDQGSSELDNIE-ELIDLDLQNSQLPQLFPDTRPGAP 830
Db 703 --SNKLIKLRQLEYEQAFQDLSGDDPPGQSTSHLMKMKMLRGSGCP-LMPD----- 753
QY 831 AGSVQKAIINDLMQTLAENSPTVPVGAOKTALRSQSTFNNPRRGQGLRLLPNQNLPLD 890
Db 754 -----KPLSANVPNDKQTQ--NPMRG-LGH--PLRLHLPLP 783
QY 891 ITLQSPGTAGP-----FPPIRNSSPYSVIPQPMGNGQMGIGNGLGNSSTGMIGNS 943
Db 784 --QPPSAISPGENSKSRFPQCVATQYQ-----DYSISSAKHVSGM 822

```

QY 944 ASRPTWPGSEMAPOSSAVRVCATTSANRPVQ 978  
 DB 823 ASRLGPFSESYLLPBLTRYDC-----ENVNPLYG 852

## RESULT 10

US-08-785-310A-8  
 : Sequence 8, Application US/08785310A  
 : Patent No. 5840532

## GENERAL INFORMATION:

APPLICANT: McKnight, Steven L.  
 APPLICANT: Russell, David W.  
 TITLE OF INVENTION: Neuronal PAS Domain Protein  
 NUMBER OF SEQUENCES: 8  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP  
 STREET: 268 BUSH STREET, SUITE 3200  
 CITY: SAN FRANCISCO  
 STATE: CALIFORNIA  
 COUNTRY: USA  
 ZIP: 94104

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/785,310A  
 FILING DATE: 21-JAN-1997

## CLASSIFICATION: 536

## ATTORNEY/AGENT INFORMATION:

NAME: OSMAN, RICHARD A  
 REGISTRATION NUMBER: 36,627  
 REFERENCE/DOCKET NUMBER: UTSD:1226

## TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 343-4341  
 TELEFAX: (415) 343-4342

## INFORMATION FOR SEQ ID NO: 8:

## SEQUENCE CHARACTERISTICS:

LENGTH: 816 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-08-785-310A-8

Query Match 3.8%; Score 289; DB 2; Length 816;  
 Best Local Similarity 18.3%; Pred. No. 2e-12;

Matches 211; Conservative 151; Mismatches 360; Indels 432; Gaps 42;

QY 4 MGNSTDPSPAEFRKRKECPDQGPSPKRTKREKREOEKYELELAFANFNIDNF 63  
 DB 1 MDEDERAKARARANK-----SEKKRDQFVLLIKELSSMLPGTR----- 41  
 QY 64 NFKPDKALIKETVQKIOIKEOEKAAMNIDEVOKSDVSTQGVYDKDALGPMLEAL 123  
 DB 42 --KMDKTVLEKVIQGLQ--KHNEVSAQTFEICIQODWKR-----PLNSEFQIMLEAL 93  
 QY 124 DGEFFVNLGNAVENVSENYOYLRYNOEELANKSVYSILHVDHTEFYKALLPKSTVNG 183  
 DB 94 DGEVIVVTGDSITIVSDSTPLGLHLPADVMDQNLNPLPEQEHSEVYKILSSHMLVTD 153  
 QY 184 GSWSGEPFRNSHTFCRML---VKP-----LPDSEEGHON--- 217  
 DB 154 SPSPEFLKSDNLEFYCHLLRGSLNPEFTYIEYIKFVGNFRSYNNVPSPCNGCFDTLS 213  
 QY 218 OEAHQYETWQCF-----AVSQPSIKEEGBDLQSLICVARVPKMERVLPSESEPTTR 273  
 DB 214 RPHVNLGKDVCTIATVRLATPQFLKE-----MCAV-----DEPL-----EFTSR 254  
 QY 274 ODLOGITSLDTSTMRAMKPGW---EDLVRCIOKFNHOGESVSYAKRHHNEVLROG 330

DB 255 HSLWMKFLFD---HRAPPIIYLPPEVLTGSGYNYH-----DDELLARQHHLMQG 307  
 QY 331 LAFSQIYRFSLDGTLVAQOTSKILRSQTNBPOLVSLH-----MLRREGVNCVM 382  
 DB 308 KKKSCCYRFLTGQOWIMWLQTHYIYTHOMNSKPEFIVCTHVSADVVRERQDEALE 367  
 QY 383 NPDLTGQTKGKPLNLISSNSPAHQALCSGNPGQDMTLSSINFPINGPKQGMCPYRGF 442  
 DB 368 DP-----PTEAM----- 374  
 QY 443 GSGGMNHVSGMQATTPOGSNTYALKMNSPSQSSPGMNGOPTMNLSPRHMSGVAGSPRI 502  
 DB 375 -----HPSAVK-----EKDSLEPPQPFNAL-----DWGASGLPSS 405  
 QY 503 PPSQFSPAGLSHPVGCSTGNSHSTYNSLNAQALSEGHGVSIGSSIASPLKMGNL 562  
 DB 406 P-----SPSASSNS-----SHKSSHTAMS----- 424  
 QY 563 QNSPVNMNPPPLSKMGLSDKDFGLYGP-SEGTTGAQESSCHPEQKETMDPNLPAV 621  
 DB 425 -----BPTSTPKLMAENSTYALPRATLPLQELPVQ 456  
 QY 622 SSERADGOSRLDSSK-----QTKLQLLTKSDQMPPLAS-----SLSDPNKOSTGSL 672  
 DB 457 LSGAATMPALHSSASCDLTKQLLQSLPQTLQSPAPVTOFSAQFSMFOITKD----- 511  
 QY 673 PGSGSTHGSTLKEKHILHRLLDSSSPVDLAKLFAEATGKDLQSSSTAGSEVYTIQ 732  
 DB 512 -----QLEQRTRILQ-----ANIRMQ 528  
 QY 733 EPVSPKKENALLRYLLDKDTRKIGLPEITPKLERLSDKTDPASTKLIAMKEKEENS 792  
 DB 529 EELHAKIQEOLCLVQ-----DSNVQWF-LQQPAVSL 558  
 QY 793 FEFGDDPGSELNLEIILDDLQNSQLPQLPDRCPAPAGSVQKAILINDMLTANEP 852  
 DB 559 FSTQRPAAQ-----QQLQQ-----RPAAPS--QPLVNV----- 586  
 QY 853 VFPVGAQKTLARISQS-----TFNNRPQOLGRLLPNONLPL-DTILQSPGAGPF 902  
 DB 587 -PFLQGITSTQVYTNHLLRESNVISAQGPKRPSQLLPASRSLSLSPQSSSTASV 645  
 QY 903 PPIRNSPYSVIPIPGMMGNGMIGNGLGNS-STGMIGNSASRPTWPGSEMAPOSSAV 961  
 DB 646 PGLSLTITIAIPPO-----DSSQCPSPDFGHRQLRLLSOPIQPMMGSCDARQPS 700  
 QY 962 RVT-----CATTSAMNRPVGGMIRNPASIPMRPSSQPGQRTLQS 1004  
 DB 701 SRGRQVKKYAQSGVMFPSPDSHPNTSSASTPV---ILMGOAVLHPSFPASRPSPLD----- 753  
 QY 1005 QVANIGPSELEMMNGPOXSQQQAPNPQTAP-----WPESTLPIQASFASONRQPS 1058  
 DB 754 -----PQAQOQPPRYLQARTSLHSEQDLSL-----LSFFSQPOTLGY 793  
 QY 1059 SPDDLCPHPAES 1072  
 DB 794 AATQSTPQPPRPS 807

## RESULT 11

US-08-816-693A-53

: Sequence 53, Application US/08816693A

: Patent No. 5874241

## GENERAL INFORMATION:

APPLICANT: Takahashi, Joseph S

APPLICANT: Turek, Fred W

TITLE OF INVENTION: Clock Gene and Gene Product

NUMBER OF SEQUENCES: 53

CORRESPONDENCE ADDRESS:

ADDRESSEE: Dressler, Rocky, Milnamow & Katz

STREET: Two Prudential Plaza, Suite 4700

```

: CITY: Chicago
: STATE: Illinois
: COUNTRY: USA
: ZIP: 60601
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.30
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/816,693A
: FILING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: No. 5874241thrup, Thomas E
: REGISTRATION NUMBER: 33,268
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 312-616-5460
: TELEFAX: 312-616-5460
: INFORMATION FOR SEQ ID NO: 53:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 816 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
:
: US-08-816-693A-53

```

```

Query Match      3.8%; Score 288; DB 2; Length 816;
Best Local Similarity 18.6%; Pred. No. 2.3e-12;
Matches 214; Conservative 148; Mismatches 361; Indels 430; Gaps 43;

```

```

QY 4 MGEVTSRRAETRRKRCPCPOLGSPKRNTEKRNREDEKNTIEELAIIFANFNIDINF 63
DB 1 MDEDEKDRAKRASRNK-----SEKKRRDQFNVLIKELSSMLPGNIR----- 41
QY 64 NFKPDKCAILKETVQIQRKEQKAAANIDEVOKSDVSSGQGVIDKALGPMLEAL 123
DB 42 --KMDKITVLEKVIQFLQ--KHNEVSAQETICDIDQDKPFS---FLSNEETQLMLEAL 93
QY 124 DGEFFVNLQEGNVVSENVTOYLRYNOELMKNKSVSILHVGDTTEFVKLLPKSIYNG 183
DB 94 DGFVIVVTTDSIIIVSDSTPRLGHLRADVMDQNLNLFLEQHESEYVKILSHMLVTD 153
QY 184 GSNAGEPPRRKSHTEFCML--VKP-----LPDSEBECHDN--- 217
DB 154 SPSEPEFLKSDMDLEFYCHILKGLSLNPKPEPTVEYIKFVGNFRSYNNVPSPCNGCFDNTLS 213
QY 218 QEAHQKYEIMQCF---AVSOPKSIKEGEDLDQSLICVARRVPMKEBPVLPSSSEFTTR 273
DB 214 RPHVAPLKDVCFATATVATLAPFLKE-----KCYA-----DEPL---EEFTSR 254
QY 274 QDLQKQITSLDTSWRAAMKPGW--EDLVRCIOKFAHQHGESVSYAKRHHEVLRQ 330
DB 255 HSLMKELFLD---HRAPPIIGYLPFEVILGTSQVNYHI---DDLELLARHQHQLMQFG 307
QY 331 LAEQIYRFSLSDGLVAQTKSKILRSQTNREPOLVLSLH-----MLHREONVCM 382
DB 308 KKKSCCYRFILKGGQOMIWLQHYHYITTHQMNKSPFIVCTHSSVSYADVARERQELALE 367
QY 383 NPDLTGOTMGRPLNPISNSPAHQALCSGNPGQDWTLSNINFLNGPKEQMGMPMGARF 442
DB 368 DP-----PTEAM----- 374
QY 443 GSGGMNVHSGQATTPQGSNTALMKNSPQSGPGMNPQOPTSMLSPRHMSPGVAGSPRI 502
DB 375 -----HPSAVK-----EKDSSLPPQPPNAL-----DMGASGLPSS 405
QY 503 PPSQFSPAGSLHSPVGVCSSTGNSHTYNNSSIALQALSEGHSVGLSSLASPLDKMNL 562
DB 406 P-----SPSASSR-----SSKSSHT-----AMSE----- 425

```

```

QY 563 QNSPVMMNPPLSKMGSLDKDCFGLYGEPSEGTGQAESSCHPEKQETNDPNLPAYS 622
DB 426 -----PISTPKL-----MAENSTTALPRATTLPOELPVQGL 457
QY 623 SERADGQSLRHDGK---QTKLQLLTTRKSDOMEPSPLAS-----SLSDTKNDSTGSLP 673
DB 458 SQAAITPRTALHSSASODTLKQLLQSLPQTGQSPAPVPTQSAQFSMQTKID----- 511
QY 674 GSGSTHGTSLKAKKHILHRLDODSSPYDLAKTLAEATGKDLSDSSSTAPGESEVYIKOE 733
DB 512 -----OLEORTILQ-----ANIRMOE 529
QY 734 PVSPPKKEKALLRYLLDKDQTKDIGLPETTPKLERLDSKTPDASNTKLIAMTEKEENSEF 793
DB 530 ELHKIQEQLCLVQ-----DSNVQM--LQOPVAVLSLF 559
QY 794 EPGDQSELDNLEETLDDIQLNSQLPOLPPTRPAPAGASYDKQAIINDIQLMTAENSFV 853
DB 560 SSIQRPAAQ-----QQLQQ-----RPAAPS---QPOLVYN----- 586
QY 854 TPVGAQKTAALRISQ-----TFNNPREGQLRLLPNOMLPL-DITLQSPTAGAPP 903
DB 587 TPLQOIISTQVNTQHLRESNVISAQGFKPMRSSQLLPASGRSLSLPSQFSSTASYLP 646
QY 904 PIRNSSPVSVIPQPMGMGQMGIGNQGNS-STGMIGNSASRPPTMPSGEAPQSSAVR 962
DB 647 PGLSLTTTAPTPQ-----DSSQCPSPDFGHRQLRLLSQPIQIPMPSPSCDAPRSEVS 701
QY 963 VT-----CAATTSAMNRPVQGMIRNPASIPMRSPSOPGQROTLOQ 1005
DB 702 RRGQVKAQAQGVMPFSPDSHTNSASAPV---LLMGAVLHPFAPASRPPLQ----- 753
QY 1006 VNNITPSELEMMGCPQYQQAAPPNOTAP-----WPESILPTQASAFASQNRQPPGSS 1059
DB 754 -----PAQAQOQPPYLOAPTLHSEQDPDSTL---LSTFSQQPTGLGYA 794
QY 1060 PDDLCPHPAAS 1072
DB 795 ATQSTPPQPPRS 807

```

```

RESULT 12
US-08-885-291-53
: Sequence 53, Application US/08885291A
: Patent No. 6057125
: GENERAL INFORMATION:
: APPLICANT: Takahashi, Joseph S.
: APPLICANT: Turek, Fred W.
: APPLICANT: Pinto, Lawrence H.
: TITLE OF INVENTION: CLOCK GENE AND GENE PRODUCT
: FILE REFERENCE: 0290-5
: CURRENT APPLICATION NUMBER: US/08/885,291A
: EARLIER FILING DATE: 1997-06-30
: EARLIER APPLICATION NUMBER: 08/816,693
: NUMBER OF SEQ ID NOS: 55
: SOFTWARE: Patent In Ver. 2.0
: SEQ ID NO 53
: LENGTH: 816
: TYPE: PRT
: ORGANISM: Mus musculus
: US-08-885-291-53

```

```

Query Match      3.8%; Score 288; DB 3; Length 816;
Best Local Similarity 18.6%; Pred. No. 2.3e-12;
Matches 214; Conservative 148; Mismatches 361; Indels 430; Gaps 43;

```

```

QY 4 MGEVTSRRAETRRKRCPCPOLGSPKRNTEKRNREDEKNTIEELAIIFANFNIDINF 63
DB 1 MDEDEKDRAKRASRNK-----SEKKRRDQFNVLIKELSSMLPGNIR----- 41
QY 64 NFKPDKCAILKETVQIQRKEQKAAANIDEVOKSDVSSGQGVIDKALGPMLEAL 123

```

```

Db      42  --KMDKITVLEKVIQFLQ--KHNEVSAQTEICIDQDWKPS-----FLSNEEFTQLMLEAL 93
Oy      124 DGEFFVNNLEGNVFNVESENTOYLRYNOEELMKSVSYLIHVGDNHFEFVNNLPRKSYNG 183
Db      94 DGFVIYVTTDGSITIVSDSTTPRLGHLRADVMQONLNLFLPBOEHSVYKILSSHMLYTD 153
Oy      184 GSWSGEPPRRNSHTFNCRL--VKP-----LPDSEEGHND-- 217
Db      154 SPSPERFLKSDNDELEFYCHLRLGSLNPREKPEPTYEIKFVGNFRSYNNVPSGCGFDTLS 213
Oy      218 QEAHQKYTEMOCF-----AVSQPKSIKEEGEDLOSCLICVARRYRMKERYLPBSESTTR 273
Db      214 RCHVPLGKDVCFIATVRLATPQFLKE-----MCVA-----DEPL-----EESTSR 254
Oy      274 ODLOGKITSLDSTMRAMKPGW---EDLVRRCIOKFHQAQHEGSESVYAKRHHNEVLROG 330
Db      255 HSLFMKFLFD---HRAPIITIGLPEVLGTSGYNYHT---DDELLARCQOHLMOFG 307
Oy      331 LAFSQTFRSLSDGTLVAQTKSKLIRSQTTNEPOLVLSH-----MLHREQNYCVM 382
Db      308 KGKSCCYRFLTKGQOMIWLQTHYITGHQWNSKREFIVCHSVSVYADVAVERRQELALE 367
Oy      383 NPDLTGOTMGKPLNPISNSPRAHQALCSGNPGQDMTSSINPFINPKRMGMKGRFG 442
Db      368 DP-----PTAM----- 374
Oy      443 GSGGMNVHSGMOATTPGGSNYALKMNSPQSPGMNPGOPTSMLSPRHRSPPVAGSPRI 502
Db      375 -----HPSAVK-----EKDLSLEPPQFPNAL-----DMGASGLPSS 405
Oy      503 PPSQFSPAGSLHSPVGCSTGNSHSTYNSLNAQALSEGHVSLGSSLASPDLKMGNT 562
Db      406 P-----SPSASSR-----SSHKSHT-----AMSE----- 425
Oy      563 QNSPVNMNPPPLSKMGLSDKDCGGLYGEPSEGTGOAESCHPGEOKETNDNPLPAVS 622
Db      426 -----PISTPTKL-----MAENSTALPPAPLPOELPQGL 457
Oy      623 SERADGSRLLHDSKG---OTKLLOLLTTKSDOMEPSLAS-----SLSDNDSTGSLP 673
Db      458 SQALTPRLALHSSASCDLTQOLLOSLPQTGLOSPPARVQFQSGFMPOTIKD----- 511
Oy      674 GSGSTHGTSLKEKHILHRLLODSSSPVDLAKLTAETGKDLQOESSSTARGSEVITKOE 733
Db      512 -----QEOFRTRILO-----ANIRMOE 529
Oy      734 PVSPPKKNALLRYLDKDTKDIGLPEITPKLERLDSKTDPRASNTKLIAMKTEKEMSF 793
Db      530 ELKHIOQQLCIVQ-----DSNVQM--LQOPAVSLSF 559
Oy      794 EPDQPGSELDNLEELDDLQNSOLPOLPDTPRGAPAGSVDKQALINDLMQTLAENSVP 853
Db      560 SSTQIRAAQ-----QQLQO-----KRAAPS-----QQLVYN----- 586
Oy      854 TPVGAQKATLALISOS-----TENNPRGOLGRLLPNONLPL--DITLQSPGAGPRP 903
Db      587 TPLQOQITSTQVINOHLRESNVIYSAQGPKMSSSOLLPASGRSLSSLPQFSSTASVLP 646
Oy      904 PINSSSYVITPQGMNGMGNMGICNGNLGNS--STGMIGNSASAPTPPSEMAPOQSAVR 962
Db      647 PGLSLTTIATPQ-----DOSQCPSPDFGHQRLRLLSOPIQPMMPGSCDAPOPEVS 701
Oy      963 VT-----CAATTSAMNRPVQGMIRNPASISIPRRSSPGQROTLQSQ 1005
Db      702 RTGROYKYASQVWFSPDHPNTNSSASTPV---LIMGQAVLHNSFPASRSPLO----- 753
Oy      1006 VNIIGPSELMNNGGPOYSQQAAPNQTAP-----WPESILPTDAQSFASQNRQPGSS 1059
Db      754 -----PAOAOQOQPPPYLQATPLSLHSEQPDSL-----LSTFSQOQGLTGYA 794
Oy      1060 PDOLLCHPRAES 1072

```

```

Db      795 ATOSTPPQPPRPS 807

RESULT 13
US-09-496-672-53
; Sequence 53, Application US/09496672
; Patent No. 6291429
; GENERAL INFORMATION:
; APPLICANT: Takahashi, Joseph S.
; APPLICANT: Turek, Fred W.
; APPLICANT: Pinto, Lawrence H.
; TITLE OF INVENTION: CLOCK GENE AND GENE PRODUCT
; FILE REFERENCE: 0290-5
; CURRENT APPLICATION NUMBER: US/09/496,672
; CURRENT FILING DATE: 2000-02-03
; PRIOR APPLICATION NUMBER: 08/885,291
; PRIOR FILING DATE: 1997-06-30
; PRIOR APPLICATION NUMBER: 08/816,693
; PRIOR FILING DATE: 1997-03-13
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 53
; LENGTH: 816
; TYPE: PRF
; ORGANISM: Mus musculus
US-09-496-672-53

Query Match      3.8%; Score 288; DB 4; Length 816;
Best Local Similarity 18.6%; Pred. No.2.3e-12;
Matches 214; Conservative 148; Mismatches 361; Indels 430; Gaps 43;

Oy      4  MGEITSDPSRAETRRKRECPDQGPSPKRTKERNRROENKYLEELAEILFANFNIDNF 63
Db      1  MDEDEKBRARASRKNK-----SEKKRRQGFVLLIKELSSMLPGNIR----- 41
Oy      64 NEKPKDCAIKETVQKIQRIKEOEKAAANIDEVQSDVSTQGVYDKDALCPMMEAL 123
Db      42  --KMDKITVLEKVIQFLQ--KHNEVSAQTEICIDQDWKPS-----FLSNEEFTQLMLEAL 93
Oy      124 DGEFFVNNLEGNVFNVESENTOYLRYNOEELMKSVSYLIHVGDNHFEFVNNLPRKSYNG 183
Db      94 DGFVIYVTTDGSITIVSDSTTPRLGHLRADVMQONLNLFLPBOEHSVYKILSSHMLYTD 153
Oy      184 GSWSGEPPRRNSHTFNCRL--VKP-----LPDSEEGHND-- 217
Db      154 SPSPERFLKSDNDELEFYCHLRLGSLNPREKPEPTYEIKFVGNFRSYNNVPSGCGFDTLS 213
Oy      218 QEAHQKYTEMOCF-----AVSQPKSIKEEGEDLOSCLICVARRYRMKERYLPBSESTTR 273
Db      214 RCHVPLGKDVCFIATVRLATPQFLKE-----MCVA-----DEPL-----EESTSR 254
Oy      274 ODLOGKITSLDSTMRAMKPGW---EDLVRRCIOKFHQAQHEGSESVYAKRHHNEVLROG 330
Db      255 HSLFMKFLFD---HRAPIITIGLPEVLGTSGYNYHT---DDELLARCQOHLMOFG 307
Oy      331 LAFSQTFRSLSDGTLVAQTKSKLIRSQTTNEPOLVLSH-----MLHREQNYCVM 382
Db      308 KGKSCCYRFLTKGQOMIWLQTHYITGHQWNSKREFIVCHSVSVYADVAVERRQELALE 367
Oy      383 NPDLTGOTMGKPLNPISNSPRAHQALCSGNPGQDMTSSINPFINPKRMGMKGRFG 442
Db      368 DP-----PTAM----- 374
Oy      443 GSGGMNVHSGMOATTPGGSNYALKMNSPQSPGMNPGOPTSMLSPRHRSPPVAGSPRI 502
Db      375 -----HPSAVK-----EKDLSLEPPQFPNAL-----DMGASGLPSS 405
Oy      503 PPSQFSPAGSLHSPVGCSTGNSHSTYNSLNAQALSEGHVSLGSSLASPDLKMGNT 562
Db      406 P-----SPSASSR-----SSHKSHT-----AMSE----- 425
Oy      563 QNSPVNMNPPPLSKMGLSDKDCGGLYGEPSEGTGOAESCHPGEOKETNDNPLPAVS 622

```

Db 426 -----D1STPTKL-----MAENSTALPRPATLPOELPVGCL 457  
 QY 623 SERADGOSRLDSDG-----QTRKLOLLTKSDQMEPSPLAS-----SLSDTNKSTGSLP 673  
 Db 458 SQAATMPALSHSASCIDLTKQLLOSLPQTSIQSPAPVTFQSAQSFQTKD----- 511  
 QY 674 GSGSTHGTSLKEKHILHRLLODSSPYDLAKLTAATGKGLSDSSSTAGSEVITKOE 733  
 Db 512 -----OLEQRTIRILO-----ANIRMOOE 529  
 QY 734 PVSEKKENALLRYLKDQDKDIGLPEITPKLERLDSKTPASNTKLIAMKEKEEMSF 793  
 Db 530 ELKHIOEOLCLVO-----DSNVQMF-LQQAVALSLF 559  
 QY 794 EPQGPSELDNLEIIDDLONSQLPOLFPDTRGAPAGSDKQAIINDLMQLAENSPV 853  
 Db 560 SSIGRPAAQ-----QOLQO-----RPAAPS--QPOLVNV----- 586  
 QY 854 TPVGAQKTAIRISOS-----TFNPRPGOLRLPLPNONLPL-DITLQSPGAGPFP 903  
 Db 587 TPLOGQITSTVYTNQHLRESNVISAQPKMRSSQLLPAGRSLSLPSPSSSTASVLP 646  
 QY 904 PIRNSSPVYIPQGMGNOGMINOGLNS-STGMIGNSASRPTMPSGEMAPQSSAVR 962  
 Db 647 PGLSLTTLAPPO-----DSDQCPSPDFGHDROLRLLSQPIQPMFPGSCDARQSEVYS 701  
 QY 963 VT-----CAATTSAMNRPVQGMIRNPASIPMRBSSQGOQRTLOSQ 1005  
 Db 702 RTGROVYVYAOQVMPSPDSHPHTSSASTPV--LLMGAAVLHPSFSPSPLO----- 753  
 QY 1006 VMNIGSELEKMMGPOYSQOQAPPNQTAP-----WPESTLPIQASFASONRQPFSS 1059  
 Db 754 -----PAQAOQPPPIYQAPTSLSHSEQDSIL-----LSTFSQOQETTLGYA 794  
 QY 1060 PDDLCPHAAES 1072  
 Db 795 ATQSTPPQPPPS 807

RESULT 14  
 US-08-971-188-8  
 ; Sequence 8, Application US/08971188  
 ; Patent No. 6326165  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wilson, Thomas G.  
 ; APPLICANT: Heinitich, Julia N.  
 ; TITLE OF INVENTION: RECOMBINANT BHLH-PAS/JHR POLYPEPTIDE AND ITS USE TO  
 ; FILE REFERENCE: 082584/102  
 ; CURRENT APPLICATION NUMBER: US/08/971.188  
 ; EARLIER FILING DATE: 1997-11-17  
 ; EARLIER APPLICATION NUMBER: 08/843.205  
 ; NUMBER OF SEQ ID NOS: 14  
 ; SOFTWARE: Patentin Ver. 2.0  
 ; SEQ ID NO 8  
 ; LENGTH: 631  
 ; TYPE: PRT  
 ; ORGANISM: Drosophila  
 ; US-08-971-188-8

Query Match 3.8%; Score 286.5; DB 4; Length 631;  
 Best Local Similarity 22.6%; Pred. No. 2e-12;  
 Matches 158; Conservative 86; Mismatches 275; Indels 179; Gaps 35;

QY 7 NTSDPSRAETRRKKECDQLGSPKRTKRNREOKNKYIELAELEJFANFNIDIDNEFK 66  
 Db 5 NIDQKEFASKE-NHC-----EIERRRKMKATAYITELSDMV-----PTGSALARK 49  
 QY 67 PDKCAIKETVYKQIRQIKDEKKAANIDEVQKSDVSSTG---OGVIDKDALGPMMLTAL 123

Db 50 PDKLILRMAYAHMKALR-----GTGNTSSDGTYYKPSFLTQDELKHLLEFAA 96  
 QY 124 DGEFFVNVLE-GNVYFSENVTOYIRYNOEELAMKSVSILHVGDTHEFVKNLPLKSYVN 182  
 Db 97 DGEFFVNSCDGGRVLYVSDVTPVLYNTQSDWYGTSLYEHLHPDREKIRQLSTQESON 156  
 QY 183 GGS-----WSGEPFRNSHT-----FNCRMLVPLPDSPEEGHNOEAOH----- 222  
 Db 157 AGRILDLKSGTYVKEGHOSMRLSMGARRGFTICARGVANPEBMSVGLNLRKQNSLGP 216  
 QY 223 -----KYETMOC--FAVSQPKS-----IKEEGEDLOS--CLICVAR-----RVPMKERP 262  
 Db 217 SRDGTNVAIVHCTGYIKWMPPTDMPNHNMERDVMDSHCICLVAIGRLQYTSRANDMS 276  
 QY 263 VLPSESEFTTRQDLOGKITSIDTSTMRAMKPGMEDLYRCLQKFAHQHEESYAKRH 322  
 Db 277 GSNQSEFITHRAMDKETFDORVNLIGYTPPELLKICYDFHR---EDOSHAKES 332  
 QY 323 HHEVLRO-GLAFSQIYRSLSDGT-LVAQTKSKLIRSQTNEDPOLVLSLHMLHREQMVC 380  
 Db 333 FDQYLVKQGMFSLIYRARKANSEYVWLRTQAYAFNLPTDEVEYI-----VC 381  
 QY 381 VMNPDLTGQTM-GKPLNDISSNSP-----AHQALCSGNFGQDMLSSNINPEIN 428  
 Db 382 T-----NSSGKTWNGAPLDAAAAHTRPOVQOQOQOQEQHVYQAAAGVDYA----- 428  
 QY 429 GKPEOMGPKGRFGSG-GMNVHSGMQATTPQGSVYALKMNSPSSSGMMPGQPTS-ML 486  
 Db 429 --RREL-TPVGSATNDGMYQTHMLAQAPPO-----QOQOQOQREGSAQTTPVGYTY 478  
 QY 487 SPRHSPGVAG--SP--RIPPSQFSPAGSLHSPVGVCSSTGNSHSYNSSLNL----- 537  
 Db 479 DTH--SPYSAGGTSPLAKIKSGTSP-----TPV-----APNSMALRPQOQ 519  
 QY 538 ---QALSEGHGVSLGSSLASP-----DLKMGN-----LQNSPVMMNPPLSKM 577  
 Db 520 QOQOQVTEGYOYQOTSPARSPSGPTYQLSAGNGRQOAPGAYQAGP-----PPPNAP 575  
 QY 578 GSLSDKCFGLXGPESECTQOAESCHPQOKETNDP 615  
 Db 576 GMDWQOAGG-HPHPHPTAHNPHHAGGAGAGP 612

RESULT 15  
 US-08-785-241-5  
 ; Sequence 5, Application US/08785241  
 ; Patent No. 5695963  
 ; GENERAL INFORMATION:  
 ; APPLICANT: McKnight, Steven L.  
 ; APPLICANT: Russell, David W.  
 ; TITLE OF INVENTION: Endothelial PAS Domain Protein  
 ; NUMBER OF SEQUENCES: 7  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP  
 ; STREET: 268 BUSH STREET, SUITE 3200  
 ; CITY: SAN FRANCISCO  
 ; STATE: CALIFORNIA  
 ; COUNTRY: USA  
 ; ZIP: 94104  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; OPERATING SYSTEM: IBM PC compatible  
 ; SOFTWARE: Patentin Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; FILING DATE: 17-JAN-1997  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: OSMAN, RICHARD A  
 ; REGISTRATION NUMBER: 36,627  
 ; REFERENCE/DOCKET NUMBER: UTSD:1229

TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (415) 343-4341  
 TELEFAX: (415) 343-4342  
 INFORMATION FOR SEQ ID NO: 5:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 875 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: peptide  
 US-08-785-241-5

## Query Match

3.7%; Score 283.5; DB 1; Length 875;  
 Best Local Similarity 19.3%; Pred. No. 5.5e-12;

Matches 207; Conservative 152; Mismatches 381; Indels 335; Gaps 43;

```

QY 17 RRRKCPDQLGSPKRNTERKRNREQENKYTELA-ELIFANENDIDNFKPDKCAILKE 75
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 5 KKKRRSSSELRKESRDACRKRSETEVEYELAHLELPHP-----SVSSHLDKASIMRL 59
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 76 TYKQIRQKEQKAAANIDEVOKSDVSSGCGVIDKDALGPMMLALGSEFFVNLGN 135
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 60 AIFLTHLHLLSSVCSSESEAD-----QQMDNLYLALBGLIAVVTQDGD 107
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 136 VFEVSENVQYLRNOELNANKSVSILHYGDHTEFEVKNLPSKSYNGSGWSGEPFRNS 195
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 108 MIFLSNISKFMGLQVELGHSIFDFTHCDEHEIRNLTN-----GSGFG----- 156
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 196 HTFNCMLYKPLPDSSEEGHNOAHQKYTEMOCFAVSQPKSIKEG----- 242
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 157 -----KKSIDSTERDFPRMKTVTNNGRTVNLKSATWKSVLHCTGOV 200
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 243 -----EDQSCILCYARVPMKERPLP--SSSEFTTRODLOGKITSLDT 285
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 201 RYNNCPHSSILGSGKEPLSLCIIWCEPQHSMDIPLDSTKTFISRHMDMKFTYCD 260
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 286 STWRAAMKPGW--EDLVRCIQFHAQHEGESVYAKRHHEVYLRGLAFLSQTIFRESLD 343
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 261 RILELI---GYHPELIGRAVEFYHALDSEMT---KSHQNLCTKGQVVSQGYRMLAKH 314
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 344 GTIYAQTKSLKIRSGTNEPOLVISLH--MLHREQNVCMNDLIGQTMGRP-----LNP 397
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 315 GGYVWLETOGTIVYPRNLQPCICMVCNVYLSIEKNDVFSMDQT--ESLTKPLMANNS 373
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 398 ISSNSPAHQALCSGNPGODMTLSSNIFPINCGRKQMGPMGRFGSGGMNHSQMAQT 457
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 374 IIPSS-----DDVAVTEKSNYLTFTKLE-----PEELAQAPY 467
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 458 PGCSNVALKMNPSQSSPG-----MNPQOPTISMLSPRHRMSPGVAGSPRIIPSPQSPAG 511
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 408 PGDAIISLFGSONFDEPSAYGKAILPQGPQWVSGLRSH-----SAQSESG 453
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 512 SLHS--PVGCSSTGNSHSTNSSLNALQALSEHGVSLSGLSLAPDLKMNGLNNSVYNN 570
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 454 SLPAFTVPOADTGTNTPSASS-----SSCSTPSSPEYYSLSLN--PLKTE 499
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 571 PPLISKMSLSDSKDCTGLYGEPESETTGAQESCHPEQKETNDPMNPRAVSSERADGOS 630
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 500 --VIEKLFAMD-----EPDPGSTQTDPS--ELDETLAPYIP-----MDGED 539
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 631 RLHDSGQTKLQL-----LTKSDQNEPSPPLASSLDTNKDSTGLPSGSGSTHGT 662
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 540 -----FQLSPICEPEPLMPESPPTPOHCFSTWMSIFQPLT-----PGATHGPF 583
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 683 LKMK-----HKLHLRLDSSSPVULAKITAAT-----GK----- 713
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 584 FLDKYQOLESRTSESEHWMSSIFPDAGSGSLSPCCGASPLPSMGMSTWMPDP 643
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 714 -----DLQESSSTAP--GS--EVTIKQEP--VSPKKKENALLRLLDKDQTKDI 757
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 644 PLHFGPTKMPVQGQASLSCALPVGSMQLELPAPLHVSMFKMRA-----KDF 692
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

QY 758 GLPEITPKLERLDSKTDPAASNKLIAKTE---KEEMSEPGDQSGSELNLEEILDIQ 814
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 693 G-----ARGPYKMSPA-----MIALSNKLKLRQLEVE----- 720
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 815 NSQLPOLPPTRGAPAGAVSDKAIINDLQLAENSPTYPVQAOKTALRISGFENNR 874
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 721 ----EQAFQDTSGDPPGTSSSHLMKRMKSLMGTCPLMP----- 757
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 875 PGQLGRLLPNQNLPLDITLQSPGAGFPPIRNSPYSVLPQGMNGMIGNOG----- 930
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 758 ----DKTISANAPDEFQKSMRGLG--QPLRLLP-----QPPSTRSSGEMNAKTGFPPQ 807
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 931 -----NLGNSSTGMIGNSASRPTMPSGEWAPOSSAVRYTCATTSAMNRPVQ 978
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 808 CVASQFQDGPQPCAGVQSVASRLGPSPFEPYLLPELTRYDC-----EVNVVPVG 857
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

## RESULT 16

US-08-227-536-2

; Sequence 2, Application US/08227536

; Patent No. 5658784

; GENERAL INFORMATION:

; APPLICANT: Eckner, Richard

; APPLICANT: Ewen, Mark

; APPLICANT: Livingston, David

; TITLE OF INVENTION: NUCLEIC ACID, ENCODING TRANSCRIPTION

; TITLE OF INVENTION: FACTOR P300 AND USES OF P300

; NUMBER OF SEQUENCES: 13

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Weingarten, Schurgin, Gagnebin & Hayes

; STREET: Ten Post Office Square

; CITY: Boston

; STATE: MA

; COUNTRY: US

; ZIP: 02109

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/227,536

; FILING DATE: 14-APR-1994

; CLASSIFICATION: 436

; ATTORNEY/AGENT INFORMATION:

; NAME: Williams Ph.D., Kathleen A.

; REGISTRATION NUMBER: 34,380

; REFERENCE/DOCKET NUMBER: DPCI-308XX

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (617) 542-2290

; TELEFAX: (617) 451-0313

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2414 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-08-227-536-2

## Query Match

3.6%; Score 277; DB 1; Length 2414;

Best Local Similarity 18.9%; Pred. No. 8.2e-11;

Matches 222; Conservative 117; Mismatches 363; Indels 474; Gaps 50;

```

QY 381 VNPDLTGQTMGRPLMPI-----SSNSPAHQALCSGNPGODMTLSSNINPING 429
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 47 LINSTELGLTNGDINOLOTSIGMOVDAASKRHQISELLRSGS-----SPYLANGVG 99
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 430 PKQGMGMPGRFEGSGGMNHSVSGMAQTTPQGSNYVALKMNSPQSSPGM-----NPGQ 481
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 100 PGQVMA-----SQAGQSSPGLGLINSVMSPT 127
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 482 PTISMLSPRHRMSPGVAGSPRIIPSPQSPAGSLHSPV---GVCSSTGNSHSTNSSLN--AL 537
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```



```

Db 128 QAGLTSP--NMGMGTSG--PNOGPTQ--STGMNNSPVNOFAMGMNTG-----TNAGNPGM 177
Oy 538 QALSGCHV-----SLGSSLASPDLK-----MGNLONSPPVNMNPPLS----- 575
Db 178 LAAGNGOGIMPNQVNGNGSGAGRGRODMQYRPNMGMSAGNLLTEPLQOQSPOMGOTGLR 237
Oy 576 -----KMGSLDSKDCFGLYGEPSEGTGQAESCHPEGEKETN---DPNLPPAVSSERA- 626
Db 238 GPQPLKMGMMNPN---PYGSPYTONPGQOIGASGLGIQYKTYLSNNLSFPADKKAV 294
Oy 627 --DQGRHLDSKQGTQLQLLTTKSDOMP---SPLASSLSTNNDSTGSLPGSGSTHGT 681
Db 295 PGGGMPNM---GQGPAPQV-----OOPGLVTPVAQ-----GMGSGAHT 329
Oy 682 SLKEKHILHRL-----QDSSSPVD-----LAKLTPAATGKD-- 714
Db 330 ADPEKRKLIOOVLVLLAHAKCQREBQAGEVRQCNLPHCRTMKVNLNMTFCOSGSKSQ 389
Oy 715 -----LSQESSSTAPGSEVTIKOEPVSPKKENALLRYL---LDKDDTKDIGLPE 761
Db 390 VAHCASSRQIISHMKNCNTRHDCPVCLPLKNADKNQOPILTGARVGLGNPSLSYG--QO 448
Oy 762 ITPKLERLDSKTDPAASNTKLA-----MKTE-----KEMSFEPGDPGS--ELD 804
Db 449 SAPNLSTV--SQIDPSSIERAYALGLPYOVNQMPQPOVQAKNQOQOGSPQGMRPMS 507
Oy 805 NLEELDLQONS---QLQOLFPTDTPRGAAGVSKQAIINDLMQITAEKSPYTPVGAQKT 861
Db 508 NMSASPMGVNGVGVOTPSLLSDM-----LHSAINSQNPMSSENAVSPLGPMPT 558
Oy 862 AL-----RISOSTFNNPRGOLG----- 879
Db 559 AAQPSSTGIRKQMHEDITODLRNLVHKLVQAIFFTPPDALKDKRMENLVAYARKVED 618
Oy 880 -----RLLPNOLPLDITLQO-----PTGAGPFPPIRNSSPYVIP 915
Db 619 MYESANRBEYHLLAEKIYKIQELEEKRTRLQKQNNLPRNAGAVPVSNNPGNMGQ 678
Oy 916 QPGMNGNMGIGNNGNLSSTGMIGNSRPTMPSGEMAPSSAVRYTCAATTSAMNRP 975
Db 679 QPGMTSN-----GLPDPS--MIRGSVPNQMP----- 704
Oy 976 VOGMIRBPASIPMRPSSQPGQROTLOSQVNMNIGPSELEMMGGPOYSQOQAPNQTAP 1035
Db 705 -----KITPOSGLNQFG-----QMSMAQPPYVROTTPPLQ--- 734
Oy 1036 WPESTILPIDQASFASQNRQPFSSPDLLCPRPAESPSDEGALLDOLYLALRNFGLLE 1095
Db 735 -----HHGQLAQOPALNPR--MGYGRMQQPSNNGOQLPQ----- 767
Oy 1096 IDRALGIELVQSQAIVPEDEPSSQDSDSNIMEQKAPVFFPOYASQAQMAQGS---YSPMQ 1152
Db 768 -----TOPPSQGMNVTNIPRLAPSSQGAQVSOAQMSSSCPVNSPIM 808
Oy 1153 DPNHETMGQRPSTYATLRMQPRGRLPRTGLVQONQPNQLRQLQHRQLAQOQNRQPLNNQJSN 1212
Db 809 PP-----GSQSHHCPOLPOLALH-----QNSPSP----- 834
Oy 1213 VSNVNLTRPGVPTQAPINAGMLAOREILNQHRLROROMHQOQOVQORTLLMRGQGLNM 1272
Db 835 -----VPSKTP-----TPH-----HPPPSI-----GAQO 853
Oy 1273 TPSSWABSGMPTATMSNPRIPOANA-----QDFPPPNNGISQODPDGFTGATTPOSPLMS 1327
Db 854 PPATTIPAPVPTPPAMPGPQSOALHPPRPQTPPTTQOLPOQVOPSLPAPASADPOQO 913
Oy 1328 PRMAHT-----QSPMMQSOQANPAYOAPSDING 1355
Db 914 PRSQOSTAASVTPNAPILPPOPATPLOSAPAVSIEG 949

```

```

PCT-US95-04682-2
: Sequence 2, Application PC/TUS9504682
: GENERAL INFORMATION:
: APPLICANT:
: TITLE OF INVENTION: NUCLEIC ACID ENCODING TRANSCRIPTON
: TITLE OF INVENTION: FACTOR P300 AND USES OF P300
: NUMBER OF SEQUENCES: 13
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Weingarten, Schurgin, Gagnebin & Hayes
: STREET: Ten Post Office Square
: CITY: Boston
: STATE: MA
: COUNTRY: US
: ZIP: 02109
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US95/04682
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/227,536
: FILING DATE: 14-April-1994
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Holliday C. Heine, Ph.D.
: REGISTRATION NUMBER: 34,346
: REFERENCE/DOCKET NUMBER: DPCI-308Xq999
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617) 542-2290
: TELEFAX: (617) 451-0313
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2414 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: PCT-US95-04682-2

Query Match 3 6%; Score 277; DB 5; Length 2414;
Best Local Similarity 18.9%; Pred. No. 8.2e-11;
Matches 222; Conservative 117; Mismatches 363; Indels 474; Gaps 50;

Oy 381 VANNPLTGQTMKPLNPI-----SSNSPAHQALCSGPNQODMTLSSNINFPNG 429
Db 47 LINSTELGITNGDINDINOQTSIGWYODASAKIKOLSELLRSGS-----SPNLNMGVGG 99
Oy 430 PREQMGMPGRFSGSGGMNHSVGMQATTPQGSNVYALKMNSPSSQSPGM-----NPGQ 481
Db 100 PQQVNA-----SOAQOSSPGLGLINSVKSVPWT 127
Oy 482 PTSMLSPPRRHMSPGVAGSRIPRPSQSPAGSLHSPV---GYCSPNGNSHSTYNSSLN-AL 537
Db 128 QAGLTSP--NMGMGTSG--PNOGPTQ--STGMNNSPVNOFAMGMNTG-----TNAGNPGM 177
Oy 538 QALSGCHV-----SLGSSLASPDLK-----MGNLONSPPVNMNPPLS----- 575
Db 178 LAAGNGOGIMPNQVNGNGSGAGRGRODMQYRPNMGMSAGNLLTEPLQOQSPOMGOTGLR 237
Oy 576 -----KMGSLDSKDCFGLYGEPSEGTGQAESCHPEGEKETN---DPNLPPAVSSERA- 626
Db 238 GPQPLKMGMMNPN---PYGSPYTONPGQOIGASGLGIQYKTYLSNNLSFPADKKAV 294
Oy 627 --DQGRHLDSKQGTQLQLLTTKSDOMP---SPLASSLSTNNDSTGSLPGSGSTHGT 681
Db 295 PGGGMPNM---GQGPAPQV-----OOPGLVTPVAQ-----GMGSGAHT 329
Oy 682 SLKEKHILHRL-----QDSSSPVD-----LAKLTPAATGKD-- 714

```

```

Db 330 ADEEKRLLIOOVLVLLHAHKCORREOANGEBVOCNLPHCRTKKNVLMNHTHCOSGKSCQ 389
QY 715 -----LSQESSSTAGSEVITIKOEVPSPKKENALLRYL---LDKDDTKDIGLPE 761
Db 390 VAHCASSRQIISIMKNTTRDCVYCLPLKNAQKRNQOPLLTGAPVGLSPSSIGV-QQ 448
QY 762 ITPKLERLDSKTPDASNTKLLIA-----MKE-----KEEMSEFPGDQGS--ELD 804
Db 449 SARNLSTV-SQIDPSSIERAYALGLPYQVQNPOTPOVQAKNOQNOQPOGSGMRPMS 507
QY 805 NLEBIIIDDLONS---QLPOLFPTTRGAPAGSDKAIINDLMQLTARENPPVYVQKQT 861
Db 508 NMSASPGVGVGVQVTPSLSDSM-----LHSAINSQNPMMSENASVSLGMP 558
QY 862 AL-----RISQSTFNNRPGOLG-----879
Db 559 AARPSTTGIRKQNHEDITQOLRNHLVHKLVOALFPTTDPALALDRMENLVAAKRYEGD 618
QY 880 -----RLPNQNLPLDITLOS-----PTGAPRPPIRNSSPYVIP 915
Db 619 MYESANRAEYHLLAKIKYIKOLEEKRRTRLOKONMLPNAAGVPMNPGPMNGOP 678
QY 916 QPQMGNOGIGNOGNLNGNSTGMIGNSASRPTMPSEMAPOSSAVYTCATTSANRP 975
Db 679 QPQMTSN-----GRLPDS--MIRGSVPNQMP-----704
QY 976 VQGGIMNPAASIPMRSSQGOQROTLOSQVMNIGPSELENNMGPOYSQOQAPNPOTAP 1035
Db 705 -----RITPQSLNPG-----QMSAQAIPPIVROTPPLQ---734
QY 1036 WPESILPIDQASFRQNRQPGSSPDDLCPHRAESPSDEGALLDQLYALRNFDGLEE 1095
Db 735 -----HHQOLAQGALNPP--MGYGRMQQPSNOGQFLPQ-----767
QY 1096 IDRALGIPELVSGQADVPQSSQSDSNIMLEQKAPYFQQAQYSAQMAQGS---VSPMQ 1152
Db 768 -----TQPSQGMNVTNIPLAPSSGAPVSAQMSSSSCPVSPI 808
QY 1153 DPNEHTMGQRPXYATLMOQRPGLRPTGLVQONQPNOLRLQHLRLQAQNRQPLMNQISN 1212
Db 809 PR-----GSGGSHHCQQLQPALH-----QNSPSP-----834
QY 1213 VSNVNTLIRGCVPTQAPINQMLAQORELILNHLRQMHQOOQVOQRTLMRGGLM 1272
Db 835 -----VPSKRP-----TPH-----HTPPI-----GAQ 853
QY 1273 TPSVAPSGMPATMSNRIPQANA-----QOFPPRYGISQODPQFGTATTPQSPILMS 1327
Db 854 PRATTIDAVPTPPAMPQFQSOALHPPRQOTPPPTTQLPQOVQPSLPAPASADOPQOQ 913
QY 1328 PRMAHT-----QSPMQOQSANPAYQAPSDING 1355
Db 914 PRSQQSTAAVPTPPMAPLPPQATPLSQAPVSIQ 949

```

```

RESULT 18
US-08-785-241-6
; Sequence 6, Application us/08785241
; Patent No. 5695963
; GENERAL INFORMATION:
; APPLICANT: McKnight, Steven L.
; APPLICANT: Tian, Hu1
; APPLICANT: Russell, David W.
; TITLE OF INVENTION: Endothelial PAS Domain Protein
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 268 BUSH STREET, SUITE 3200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:

```

```

; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/785,241
; FILING DATE: 17-JAN-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A.
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UTSD:1229
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 343-4341
; TELEFAX: (415) 343-4342
; INFORMATION FOR SEQ. ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 826 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-785-241-6

```

```

Query Match 3.5%; Score 269.5; DB 1; Length 826;
Best Local Similarity 19.3%; Pred. No. 5,2e-11;
Matches 186; Conservative 150; Mismatches 337; Indels 289; Gaps 41;

```

```

QY 1 MSGKENTSPSRAETRRKRECPDOLGSPKRTNTEKRNQOEKKYIELEALIFANFNDI 60
Db 1 MEGAG-GANDKKRISSEKRE-----KSRDAASRRKRESEVEYELAHQPLPHNS 51
QY 61 DNFNFKPKCAIKETVYKQIRKDEQKAANANIDEVQKSDVSTGQYIDKALCPMMI 120
Db 52 SHL-----DKASVRLITSTYR-VKRLLDAGDDIEDDKAQMNC-----FYL 93
QY 121 EALDGEFFVNLGVNVESENVTQYLRYNQOELEMKSVYSILHVDHTEFVKNLPRKST 180
Db 94 KALDGEFVMTLDDGDMYISDNVKNYKWLQFELTGHSVDFTHPDHDEMRMLTHR-- 151
QY 181 VNGSMGGERPRNRNHTFNCRMLVYKRLPSEEGHONQEHQKXYEMQCAVSPRSIKE 240
Db 152 -NG-----LVK-----KGRQONTQSRSEFLMKCTYLSRGRTMT 184
QY 241 EG-----EDLQSLICVARRVPMKERVLP--SSESEFTT 272
Db 185 KSAWKVYLHCTGHIHYDITNSNQPCYKKRPMTCLVLICEPIPHNSNIEIPDSTFELS 244
QY 273 RQDLQKITSLDTS-TWRAAMKPGMEDLVRCIOKFAHOGESVSYAKRHHHEVLROGL 331
Db 245 RHLSDMKFSYCDERITELMGEPR--EELLGRSIYEY--HALDS-DHLTKTHHDMFTKQ 299
QY 332 AFQOIYRFSLSDGTIVAQAQKSKLINSQITNEPOLVLSLH-----MLHRQNV 380
Db 300 VTTGQYRMLAKRGQYVWVEQATVYNTKNSQPCQICVAVYVSGIIGHDLITSLQDTEC 359
QY 381 VMNP-DLTGQTMKPLNPISNSPAH-----QALCSGNPGODMTLSSNINFPINGP 430
Db 360 VLPKRVESDMKMTQLFTKVESEDTSLFDLKKEPDALITLLAAAGGTITIS-IDF-----413
QY 431 KEQMGMPGRFGSGGNHVSQMAQATTPQGSNAYALKMNSPQSGPCGNPGOPTSMISPRH 490
Db 414 -----GSNDTETDDQOLEVPLVYN--DVMLSPSNE 441
QY 491 RM-SPGVAGSPRIPTPSFAGSLHSPVGVCSSTGNSHSTNSIALQALS-EGHCYSL 548
Db 442 KIONINLAMP-----LFTAETPKVL-----RSSADPALNOEVALKLEPPESLT 485
QY 549 GSSLASPDLLKGNLONSVMNPPPLSKMGLSDKDCFGLYGPESEGTQGAESSCHPGE 608
Db 486 ELSTMTQIQ-----DQTP-----SPDSQSTNQ--SSRPPNS 515

```

QY 609 QKE-----TNDPNLPRAVSSERADGQSLHDSKQTKL-LQLLT-----T 647  
 Db 516 PSEYFYVDSDVWNEFKLELVEKLEFAEDTEAKNPFSTQDTDLLEMLAVYIPMDDFOLR 575  
 QY 648 KSDQMEPPLASSLSDTKNDSTGSL-----PGSGTHGSLKXKHLH----- 691  
 Db 576 SPDOL--SPESSSASPESASPOSTVTVFOQTQIOEPTANATTATTDLTETVTKDRME 633  
 QY 692 --RLDSSSPVDLAKLTAEAT-----GRDLSQESSSTAGS----- 726  
 Db 634 DIKILASPSPIHKEITTSATSPYROTQSTRTASPNRAGKVIEOTEKSHRSPNVL5V 693  
 QY 727 ---EVITIKQEPVSPK--KRENALLRYLLDKDDT--KDIGLEIPPKLERLDSKTDPSAN 778  
 Db 694 ALSQRTVPEEELNPKILALQNAQRKRMEHDSLEQAVIGITL-----LQOPDDHAAT 747  
 QY 779 TKLIMKTEKEMSFEPGQPSSELDNLEIIDL-----ONSQLPQLFP-DTRG 828  
 Db 748 TSL---SWKRVKGCSSBQNGMEQKTIILIPSLACRLGQSMDESGLPOLTSYDCEVN 803  
 QY 829 AP 830  
 Db 804 AP 805

RESULT 19  
 US-08-480-473B-2  
 : Sequence 2, Application US/08480473B  
 : Patent No. 5882914  
 : GENERAL INFORMATION:  
 : APPLICANT: Semenza, Gregg L.  
 : TITLE OF INVENTION: HYPOXIA INDUCIBLE FACTOR-1 AND METHOD OF USE  
 : NUMBER OF SEQUENCES: 64  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: Fish & Richardson P.C.  
 : STREET: 4225 Executive Square, Suite 1400  
 : CITY: La Jolla  
 : STATE: CA  
 : COUNTRY: USA  
 : ZIP: 92037  
 : COMPUTER READABLE FORM:  
 : MEDIUM TYPE: Floppy disk  
 : OPERATING SYSTEM: IBM PC compatible  
 : SOFTWARE: Patent Release #1.0, Version #1.30  
 : CURRENT APPLICATION DATA:  
 : APPLICATION NUMBER: US/08/480,473B  
 : FILING DATE: 06-JUN-1995  
 : CLASSIFICATION: 514  
 : ATTORNEY/AGENT INFORMATION:  
 : NAME: Halle, Lisa A.  
 : REGISTRATION NUMBER: 38,347  
 : REFERENCE/DOCKET NUMBER: 07265/053001  
 : TELECOMMUNICATION INFORMATION:  
 : TELEPHONE: 619/678-5070  
 : TELEFAX: 619/678-5099  
 : INFORMATION FOR SEQ ID NO: 2:  
 : SEQUENCE CHARACTERISTICS:  
 : LENGTH: 826 amino acids  
 : TYPE: amino acid  
 : TOPOLOGY: linear  
 : MOLECULE TYPE: protein  
 : US-08-480-473B-2

Query Match 3.5%, Score 269.5; DB 2; Length 826;  
 Best Local Similarity 19.3%, Pred. No. 5.2e-11;  
 Matches 186; Conservative 150; Mismatches 337; Indels 289; Gaps 41;

QY 1 MSGMGENTSDPSRAETKRKEPPDGLGSPKRNTEKRNREQENKIELELAELIFANFDI 60  
 Db 1 MEGAG-GANDKKKISSERRK-----KSRDAARSRRSKESSEVYELAHQLPLPHNVS 51

QY 61 DNFNRPKCAILKETVKQIROIKDOEKAAANIDEVOKSDVSTGQGVIDKDALGPMML 120  
 Db 52 SHL---DKASVWRILISYLR-VRKLLADGDDIDDDMKQAQNC-----FYL 93  
 QY 121 EALDGEFFVYNLEGNVVESEVNTQYLRNOBELANKSVSYSLHNGDTEFYKNLPSI 180  
 Db 94 KALDGEVAVLTDGDMITVSDVNNKMYMGLTQPELGHVSFDETHPCDHEKREMLTHR-- 151  
 QY 181 VNGSWSGEPPRRNSHTFCRLVYPLPDSSEGHNDQNAHKKYETMOCFANASQPSKIE 240  
 Db 152 -NG-----LVK-----KGAQONTORSFFLRMKCTLTSGKRTMY 184  
 QY 241 EG-----EDLOSCLICVARRVPMKEREVLP--SESEFTT 272  
 Db 185 KATMKVLTCHGTGHINHYDNTSNQPGCYKKPMTGCVLICEPIPHSPNIEIPLDCKTFLS 244  
 QY 273 RODLOGKITSLDTS--TMRAMKPGMEDLVRCIOKFAHQBESVSYAKRHHEVLRGL 331  
 Db 245 RSLDMKPFSCDERITELMGEYR--EELIGRSIYEY--HALDS-DHLTKTHHDFTKQ 299  
 QY 332 AFSQIYRFSLSQDTLYAAQTKSLIRSQTNREPOLVISLH-----MLHREONVC 380  
 Db 300 VTTGQTRMLAKRGYVWVETQATVYNTKNSQPQCIQVNYVSGIIQHDLPISLQOTEC 359  
 QY 381 VANNP-DLTGQTMGKPLNPISNSPAH-----QALCSGNPGQDMTLSSNINPINC 430  
 Db 360 VLKPVSSDMKMTQLEFTVYESEDTSSLPDKLKEPDLALLPAGDITIS-LDF----- 413  
 QY 431 KEOMGPMKRGFGSGGMNHVSGMOATTPOGSYVALKMNPSQSSSPQMNPGQTSMLSPRH 490  
 Db 414 -----GSDTETDDQLEEVPLYN--DVMLPSPNE 441  
 QY 491 RM-SPGVAGSPRIIPSPQSPASLSIPVCGSTGNSHTSSINALQALS-BEGHVS 548  
 Db 442 KLQNLINLAMP-----LPTAETPKPL-----RSSADPALNQEVALKLEPNPSL 485  
 QY 549 GSSLASPLDKMGNLQNSPVNMPPLSKMSLDSKDCFLYGEPSGTTGALESSCHPGE 608  
 Db 486 ELSTFMPQIQ-----DQTP-----SPSDGSTRQ--SSPEPRS 515  
 QY 609 QKE-----TNDPNLPRAVSSERADGQSLHDSKQTKL-LQLLT-----T 647  
 Db 516 PSEYFYVDSDVWNEFKLELVEKLEFAEDTEAKNPFSTQDTDLLEMLAVYIPMDDFOLR 575  
 QY 648 KSDQMEPPLASSLSDTKNDSTGSL-----PGSGTHGSLKXKHLH----- 691  
 Db 576 SPDOL--SPESSSASPESASPOSTVTVFOQTQIOEPTANATTATTDLTETVTKDRME 633  
 QY 692 --RLDSSSPVDLAKLTAEAT-----GRDLSQESSSTAGS----- 726  
 Db 634 DIKILASPSPIHKEITTSATSPYROTQSTRTASPNRAGKVIEOTEKSHRSPNVL5V 693  
 QY 727 ---EVITIKQEPVSPK--KRENALLRYLLDKDDT--KDIGLEIPPKLERLDSKTDPSAN 778  
 Db 694 ALSQRTVPEEELNPKILALQNAQRKRMEHDSLEQAVIGITL-----LQOPDDHAAT 747  
 QY 779 TKLIMKTEKEMSFEPGQPSSELDNLEIIDL-----ONSQLPQLFP-DTRG 828  
 Db 748 TSL---SWKRVKGCSSBQNGMEQKTIILIPSLACRLGQSMDESGLPOLTSYDCEVN 803  
 QY 829 AP 830  
 Db 804 AP 805

RESULT 20  
 US-08-915-213-2  
 : Sequence 2, Application US/08915213  
 : Patent No. 6020462  
 : GENERAL INFORMATION:  
 : APPLICANT: Semenza, Gregg L.  
 : TITLE OF INVENTION: HYPOXIA INDUCIBLE FACTOR-1 AND METHOD OF USE  
 : NUMBER OF SEQUENCES: 64

```

CORRESPONDENCE ADDRESS:
ADDRESS: Fish & Richardson P.C.
STREET: 4225 Executive Square, Suite 1400
City: La Jolla
STATE: CA
COUNTRY: USA
ZIP: 92037

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/915,213
FILING DATE: 20-AUG-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/480,473
FILING DATE: 06-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Halle, Lisa A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 07265/053001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619/678-5070
TELEFAX: 619/678-5099
INFORMATION FOR SEQ. ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 826 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-915-213-2

```

Query Match 3.5%; Score 269.5; DB 3; Length 826;  
 Best Local Similarity 19.3%; Pred. No. 5,2e-11;  
 Matches 186; Conservative 150; Mismatches 337; Indels 289; Gaps 41;

```

QY 1 MSGGENTSPSRAETRRKCEPDQIGPSPKRNTKRNROENKYEIELAELEAFANNDI 60
DB 1 MEGAG-GANDKKKISSRRE-----KSRDAARRKSESEVEYELAHOLPLPHNVS 51
QY 61 DNFNFKPKCAILKETVYKQIRKEDOKAANAANIDEVOKSDVSTGGVIDKDALGPMML 120
DB 52 SHL-----DKASVWRLLTSTYLR-VKRLDAGDLDIEDMKAKQMC-----FYL 93
QY 121 EALDGEFFVNLGNVFESENTOYLRYNOEELMKNVSYLAVGDHTEFVKNLLPKST 180
DB 94 KALDGFVWVLTDDGMILYISDNVKNKMGILQFELTGHVSDFTHPCDHEMRMLTHR-- 151
QY 181 VNGGSMGEPFRNRSHTFNCRMLVYKPLPDESEEGHNOEAKYETMOCFAVSQPSIKE 240
DB 152 -NG-----LVK-----KGEONTORSFLLMKCTLTLSRGRTMNI 184
QY 241 EG-----BDLOSCLICVARVPYMKERPYLP--SSESFTT 272
DB 185 KSAFWKVLHCTGHIVDTNSNOQGYKKRPMTCVLVCEPIPHSPENIELPSKFTLS 244
QY 273 RODLOGITSLDTS-TYRAAMKPEWEDLVYRCIOKFAHOGESVSYAKRHHHEVLRQGL 331
DB 245 RHSLDMKFSYCDERITELMGEYR--EELIGRSIYEY--HALDS-DHILTKHDMFTKGO 299
QY 332 AFSQIYFSLSDGTLVAAGKRSKLIRSQTNPEQLVLSLH-----MLHREQNV 380
DB 300 VTTGVGVMLAKRGQYVVEVQATVYTNKNSQPCQICVAVVSGIIOHDLITSLQOTEC 359
QY 381 VAMP-DLTGOTMGKPLNPISNSPAH-----QALCSGNPGODTSLSSINIFNPG 430
DB 360 VLKPVESDDMKQTLFTKVSSEDTSSLFDKLKKEPKDALITLAPAGTTIIS-LDF----- 413
QY 431 KEOMGMPMGRRFGSGGNHVSIGMOATTPGGSNVALKKNSSQSSPGMNPQOPTSMISPRH 490

```

```

DB 414 -----GSNDTEFDQQLLEEVPVLYN---DVMLESPNE 441
QY 491 RM-SPVAGSPRIPQSPAGSLHSPVGCSTGNSHTYNSLNAOLMS-EGHGVSL 548
DB 442 KIONINLAMP-----LPTAETPKPL-----RSSADPALNOEVALKLEPPESL 485
QY 549 GSSLASPDLMKGNLONSPVNMNPPILSKMGLSDKDCFGLYGSPSGCTGQABSSCHPG 608
DB 486 ELSTFMQIQD-----DQTP-----SPDSGSTRQ--SSPEPNS 515
QY 609 QKE-----INDNLPRAVSSERADGSRHDSKGQTKL-LQLLT-----T 647
DB 516 PSEYCFYVDSDVNEKLELVEKFAEDTEAKNPFSTQDTDLLEMLAPYPMDDDFOLR 575
QY 648 KSDOMEPSPLASLSDTNKDSGSL-----PGSGSTRGTSUKKKKILH----- 691
DB 576 SFDQL--SPLESSASBPESAPOSTYVTVQOTQIOEPTAATTTTATTDLAKTVTDRE 633
QY 692 --RLDSSSPVDLAKTLAEAT-----GKDLSSSSSTAPGS----- 726
DB 634 DILILASPEPTIHKETTSATSPYRDQSRTPASPNRAGKVLEQTEKSHRSPVLSV 693
QY 727 ----EVTIKQEPVSPK--KKNALLRYLLDKDT--KDGLPEITTKLERLDSKTPASN 778
DB 694 ALSQRTVPEEELNPKILALONARRKMEHDSLFQAVGIGTL-----LQPPDHAAT 747
QY 779 TKLIAMKTEKEEMSFEPGDPGSELNLEILLDL-----QNSQLPOLFP-DTRPG 828
DB 748 TSL-----SKRVGCKSSSEONGMEOKTIIILPSDLACRLLGQSMDESGLPOLSTYCEVN 803
QY 829 AP 830
DB 804 AP 805

```

RESULT 21  
 US-09-148-547-2  
 ; Sequence 2, Application US/09148547  
 ; Patent No. 6124131  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Semenza, Gregg L.  
 ; TITLE OF INVENTION: Hypoxia Inducible Factor-1 and Methods of Use  
 ; FILE REFERENCE: 07265/151001  
 ; CURRENT APPLICATION NUMBER: US/09/148,547  
 ; CURRENT FILING DATE: 1998-08-25  
 ; NUMBER OF SEQ. ID NOS: 2  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ. ID NO 2  
 ; LENGTH: 826  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-09-148-547-2

Query Match 3.5%; Score 269.5; DB 3; Length 826;  
 Best Local Similarity 19.3%; Pred. No. 5,2e-11;  
 Matches 186; Conservative 150; Mismatches 337; Indels 289; Gaps 41;

```

QY 1 MSGGENTSPSRAETRRKCEPDQIGPSPKRNTKRNROENKYEIELAELEAFANNDI 60
DB 1 MEGAG-GANDKKKISSRRE-----KSRDAARRKSESEVEYELAHOLPLPHNVS 51
QY 61 DNFNFKPKCAILKETVYKQIRKEDOKAANAANIDEVOKSDVSTGGVIDKDALGPMML 120
DB 52 SHL-----DKASVWRLLTSTYLR-VKRLDAGDLDIEDMKAKQMC-----FYL 93
QY 121 EALDGEFFVNLGNVFESENTOYLRYNOEELMKNVSYLAVGDHTEFVKNLLPKST 180
DB 94 KALDGFVWVLTDDGMILYISDNVKNKMGILQFELTGHVSDFTHPCDHEMRMLTHR-- 151
QY 181 VNGGSMGEPFRNRSHTFNCRMLVYKPLPDESEEGHNOEAKYETMOCFAVSQPSIKE 240
DB 152 -NG-----LVK-----KGEONTORSFLLMKCTLTLSRGRTMNI 184

```

```

OY 241 EG-----EDLOSLICVARRVPMKRPVLP-SSESEFTT 272
DB 185 KSAWKVILHCTGHIHVYDTSNQPGYKKRPMTCVLICERIPRPSNIEIPDLSKTFLS 244
OY 273 RODLOGKITSLDTS-TYRAAKPGMEDLVRRICIOFHNOHGESVYAKRRHHNEVLROGL 331
DB 245 RHLIDMKFSYCDERTELMGYER--EELLGRSIEY--HALDS-DHLTKTHHDMFTKGO 299
OY 332 AFSQIYRESLSDGTLVAAQTKSLRSQTNPEQVLSLH-----MLHREDQVNC 380
DB 300 VTTGQYRMLAKRGYVWVETQATVLYNKNQSPQICVNVVSGIIQHDILFSLQDEEC 359
OY 381 VMNP-DLTGQTMGKPLNFISSNSPAH-----QALCSNPGODMTLSSNINPGR 430
DB 360 VLKPESSDMKMTQFLTVESEDTSSLFDKLLKRPDALTLAPAGDTIIS-LDF----- 413
OY 431 KEQMGMPGRFGSGGGMNVHSGMOATTPQGSNYALKMNSPQSSPGMNPQPTSMLSPRH 490
DB 414 -----GSDNETDQOLEEYPLVN---DVMLPSFNE 441
OY 491 RM-SPGVAGSPRIPPSQSPAGSLSPYGVCSSTGNSHSYNSSLNALQALS-EGHGVSL 548
DB 442 KLONINILAMSP-----LPTAETPKPL-----RSSADPALNOEVALKLEPNPESL 485
OY 549 GSSLASPDLKMGNLONSPVNNNPPLSKMGLSKDCGFLXGEPSEGTGGAESSCHPGE 608
DB 486 ELSTFMQIO-----DQTP-----SPSDGSTRO--SSPEPNS 515
OY 609 OKE-----TNDPNLPAVSSERADGOSRLHDSKGOTKL-LQLLT-----T 647
DB 516 PSECFYVDSUMVNEFEKLELFAEDTEAKNPFSTQDTDLDEMLAPYIPMDDDFOLR 575
OY 648 KSDMEPSLASSISDTRKDKSTGSL-----PGSGTHGSLKEKKILH----- 691
DB 576 SFDOL--SPLSSSSASPSASPOSTVTVTQOTOIEPTANMTTATTDELKTVYKDRME 633
OY 692 --RLLODSSFPVLDLAKLTAET-----GDLQSOESSSTAPGS----- 726
DB 634 DIKILIASPSPTHIHKEETTSATSSPYRDTQSRASPNRAGGVIEQTEKSHPRSPNVLS 693
OY 727 ----EVTIKOPVSPK--KRENALIRYLLDKODT--KDIGLEPTTPKLERLDSKTDPAASN 778
DB 694 ALSORTVPEELPKILALONAKRKRMEHDSLFQAVGIGTL-----LQOPDHAAT 747
OY 779 TKLIAMTEKEMSPEDQSGELDNLEITLDL-----QNSOLPOLFP-DTRPG 828
DB 748 TSL-----SMKRVKCGCKSEBQMGEMOKTILIPSLDACLKQSDMESGLPOLTYDCEVN 803
OY 829 AP 830
DB 804 AP 805

```

```

RESULT 22
US-09-235-217-2
: Sequence 2, Application US/09235217
: Patent No. 6222018
: GENERAL INFORMATION:
: APPLICANT: Semenza, Gregg L.
: TITLE OF INVENTION: HYPOXIA INDUCIBLE FACTOR-1 AND METHOD OF USE
: NUMBER OF SEQUENCES: 64
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Fish & Richardson P.C.
: STREET: 4225 Executive Square, Suite 1400
: CITY: La Jolla
: STATE: CA
: COUNTRY: USA
: ZIP: 92037
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS

```

```

: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/235,217
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/480,473
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Haile, Lisa A.
: REGISTRATION NUMBER: 38,347
: REFERENCE/DOCKET NUMBER: 07265/053001
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 619/678-5070
: TELEFAX: 619/678-5099
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 826 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-09-235-217-2

```

```

Query Match 3.5%; Score 269.5; DB 4; Length 826;
Best local similarity 19.3%; Pred. No. 5.2e-11;
Matches 186; Conservative 150; Mismatches 337; Indels 289; Gaps 41;

OY 1 MSGGENTSPPSRATRRKRCPPDGLSPKRRTEKRRNREQENKYEIELAELIFANFNDI 60
DB 1 MEGAG-GANDRKKIISERRKE-----KSRDAARSRSKSESEFYELANQLPLPHNVS 51
OY 61 DNFNFKPKCALIKETVQIQOIKOEKAAANIDDEVOKSDVSSTGQCVIDKDALCPMML 120
DB 52 SHL-----DKASVMRLTISYLR-VKRLDAGDIDEDDKAKQAMNC-----FYL 93
OY 121 EALDGEFFVNLGVNVVSENVTOYLRYNOEELMKNKSVYSIILVYGDHTEFVKNILPKSI 180
DB 94 KALDGEVAVVLLDDGDMITISDNVKNKYKGLTOPELGHSHVFPETPRCHBEKREMLTIR- 151
OY 181 VNGSWSGEPFRNSHFNCRMLVYKPLPDSSEBHDNQEAHOKYETMOCFAVSOPKSIKE 240
DB 152 -NG-----LVK-----KKEQNTQSFPLRMKCTLSRGRTWNI 184
OY 241 EG-----EDLOSLICVARRVPMKRPVLP-SSESEFTT 272
DB 185 KSAWKVILHCTGHIHVYDTSNQPGYKKRPMTCVLICERIPRPSNIEIPDLSKTFLS 244
OY 273 RODLOGKITSLDTS-TYRAAKPGMEDLVRRICIOFHNOHGESVYAKRRHHNEVLROGL 331
DB 245 RHLIDMKFSYCDERTELMGYER--EELLGRSIEY--HALDS-DHLTKTHHDMFTKGO 299
OY 332 AFSQIYRESLSDGTLVAAQTKSLRSQTNPEQVLSLH-----MLHREDQVNC 380
DB 300 VTTGQYRMLAKRGYVWVETQATVLYNKNQSPQICVNVVSGIIQHDILFSLQDEEC 359
OY 381 VMNP-DLTGQTMGKPLNFISSNSPAH-----QALCSNPGODMTLSSNINPGR 430
DB 360 VLKPESSDMKMTQFLTVESEDTSSLFDKLLKRPDALTLAPAGDTIIS-LDF----- 413
OY 431 KEQMGMPGRFGSGGGMNVHSGMOATTPQGSNYALKMNSPQSSPGMNPQPTSMLSPRH 490
DB 414 -----GSDNETDQOLEEYPLVN---DVMLPSFNE 441
OY 491 RM-SPGVAGSPRIPPSQSPAGSLSPYGVCSSTGNSHSYNSSLNALQALS-EGHGVSL 548
DB 442 KLONINILAMSP-----LPTAETPKPL-----RSSADPALNOEVALKLEPNPESL 485
OY 549 GSSLASPDLKMGNLONSPVNNNPPLSKMGLSKDCGFLXGEPSEGTGGAESSCHPGE 608
DB 486 ELSTFMQIO-----DQTP-----SPSDGSTRO--SSPEPNS 515
OY 609 OKE-----TNDPNLPAVSSERADGOSRLHDSKGOTKL-LQLLT-----T 647

```

```

Db      516 PSEYCYVDSDWVNEFKLEVERKFAEDTEAKNPFSTODDLDLEMLAPYIPMDDDFOLR 575
QY      648 KSDOMEPPPLASLSDTNKDSGSL-----PGSGTHGTSLKKEHKILH----- 691
Db      576 SFQDL--SPLSSASBPESASPOSTVTVFOQTQIOEPFANAATTITATDELKIVTDRME 633
QY      692 --RLDSSSPVDLAKLTAEAT-----GKLSOBSSSTAPGS----- 726
Db      634 DIKILASPSPTIHKETTSATSSPYRDTQSRTPASPRAGKGVIEQTEKSHRSPNVLSV 693
QY      727 ----EVTIKOEPVSPK--KKENALLRYLDKDDT--KDIGLPEITPKLERLDSKTPASN 778
Db      694 ALSQRTTVEEELNPKILALONQRRKMEHDSLFOAVGIGTL-----LQPDHHAAT 747
QY      779 TKLIAMKTEKESFEPGDPGSELNLEILDDL-----QNSQLPOLFP--DTRPG 828
Db      748 TSL-----SKRVKGCSSSQNGMEOKTIIILIPSLACLRLGQSMDESGLPOLTSYDCEVN 803
QY      829 AP 830
Db      804 AP 805

```

```

RESULT 23
PCT-US96-10251-2
; Sequence 2, Application PC/TUS9610251
; GENERAL INFORMATION:
; APPLICANT: The Johns Hopkins University School of Medicine
; TITLE OF INVENTION: HYPOXIA INDUCIBLE FACTOR-1 AND METHOD OF USE
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/10251
; FILING DATE: 06-JUN-1996
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile, Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 07265/053WO1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 826 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US96-10251-2

```

```

Query Match      3.5%; Score 269.5; DB 5; Length 826;
Best local similarity 19.3%; Pred. No.5,2e-11;
Matches 186; Conservative 150; Mismatches 337; Indels 289; Gaps 41;

```

```

QY      1 MSGMGENTSDPSAETRRRRCDDGLGSPKRNTERKRNDEQNKYIEELALIFANFNDI 60
Db      1 MEGAG--GANDKKKISSRRKE-----KSRDAARSGRSKESSEVEYELAHQPLPHNVS 51
QY      61 DNNFPRDKAIIKEYKQIROIKEDEKAAANIDEVQKSDVSTGGGVADIKALGPMML 120

```

```

Db      52 SHL-----DKASYWRLTISYLR--VRKLLDAGDLDIEDMKQAMNC-----EYL 93
QY      121 EALDGEFFVYVNEGANVFSEVENVTOYLRYNOEELMKSVYSLIHWGHTHEFYKNLPKSI 180
Db      94 KALDGFVWVLTDDGDMITISIDNVKNKMGTLQFELTGHSHSFDDTHPRDHEMKMLNHR-- 151
QY      181 VNGSNWSEPPRRNSHTFNCRLVKPLPDSEEBGHNDQANOKYETMOCFAVSQPKSKE 240
Db      152 -NG-----LVK-----KGEONQORSFFLMKCTLSRGRTNMI 184
QY      241 EG-----EDQSLICARVYMKREVRLP--SSRSFTT 272
Db      185 KSAFWKVLHCTGHIHYDTNSNQPOCGYKRPWTCLVLCERPIPHSPNIEIPDSTTELS 244
QY      273 RODLOGKITSLDPS--TMAAMKPGMEDLVRCIOKFEHNAOHESESVAYARNHNEVLOGL 331
Db      245 RHLDMKFSYCDERITELMGER--BELIGRSTIYEV--HALDS--DHLKTHNDMTKQ 299
QY      332 AFSQIYRFSISDGTIVAATQKSLIRSQTNBQVYISLH-----MLHREQNV 380
Db      300 VTTGQYRMLAKRGYVWVEFQATVIYNTKNSOPQOCIVCVNVVSGIIQHDLIFFSLQOTEC 359
QY      381 YANP--DLTGOTMCKPLNPISNSPAH-----QALCSNPGQDMTSSNINPFINP 430
Db      360 VLKPVESDMKMTQLTFTKVESEDTSLEFKLKEPDALTLLAPAGDITIS--LDF----- 413
QY      431 KEQMGPMRFGSGSGGMNVHSGMATTPQGSNVYALMKNPSPOSSPQMPNPGQPTSMLSRN 490
Db      414 -----GSDNTEHDQGLEVPRLXN--DYMLSPNME 441
QY      491 RM-SPEGVAGSPRIIPSOEPSPAGSLHSPVGCSTGNSHTYSSNLALDLS--EGHGVSL 548
Db      442 KIONINLAMP-----LPTAETPKPL-----RSSADPALNOEVALKLEPNESL 485
QY      549 GSSLASPLDKMGLQNSPVNMNPPRLSKMGSLSKDCFGLYGEPSPGCTGOAESCHPE 608
Db      486 ELSEFTMPQIQ-----DQTP-----SPDGSSTRO--SSPEPNS 515
QY      609 QKE-----TNDPLPRAVSEERADGQSRHLSKGOTKL--LDLTL-----T 647
Db      516 PSEYCYVDSDWVNEFKLEVERKFAEDTEAKNPFSTODDLDLEMLAPYIPMDDDFOLR 575
QY      648 KSDOMEPPPLASLSDTNKDSGSL-----PGSGTHGTSLKKEHKILH----- 691
Db      576 SFQDL--SPLSSASBPESASPOSTVTVFOQTQIOEPFANAATTITATDELKIVTDRME 633
QY      692 --RLDSSSPVDLAKLTAEAT-----GKLSOBSSSTAPGS----- 726
Db      634 DIKILASPSPTIHKETTSATSSPYRDTQSRTPASPRAGKGVIEQTEKSHRSPNVLSV 693
QY      727 ----EVTIKOEPVSPK--KKENALLRYLDKDDT--KDIGLPEITPKLERLDSKTPASN 778
Db      694 ALSQRTTVEEELNPKILALONQRRKMEHDSLFOAVGIGTL-----LQPDHHAAT 747
QY      779 TKLIAMKTEKESFEPGDPGSELNLEILDDL-----QNSQLPOLFP--DTRPG 828
Db      748 TSL-----SKRVKGCSSSQNGMEOKTIIILIPSLACLRLGQSMDESGLPOLTSYDCEVN 803
QY      829 AP 830
Db      804 AP 805

```

```

RESULT 24
US-08-971-188-10
; Sequence 10, Application US/08971188
; Patent No. 6326165
; GENERAL INFORMATION:
; APPLICANT: Wilson, Thomas G.
; APPLICANT: Heinrich, Julia N.
; TITLE OF INVENTION: RECOMBINANT BHLH-PAS/JHR POLYPEPTIDE AND ITS USE TO
; FILE REFERENCE: 082584/102

```

```

: CURRENT APPLICATION NUMBER: US-08/971,188
:
: CURRENT FILING DATE: 1997-11-17
:
: EARLIER APPLICATION NUMBER: 08/943,205
:
: EARLIER FILING DATE: 1997-04-14
:
: NUMBER OF SEQ ID NOS: 14
:
: SOFTWARE: PatentIn Ver. 2.0
:
: SEQ ID NO 10
:
: LENGTH: 626
:
: TYPE: PRT
:
: ORGANISM: brain and muscle ARNT-like protein a
US-08-971-188-10

```

Query Match	3.5%	Score	269	DB	4	Length	626
Best Local Similarity	24.3%	Pred. No.	3.6e-11				
Matches	142	Conservative	82	Mismatches	230	Indels	130
						Gaps	27

QY	35	EKRREDENKIIEELAEIIFANFNNDIDFNKRPDCAILKEFKVIRI0K--E0EKAAN	93
Db	81	EKRREDKMNSTIDELASLV-----PFCNMSKRDKLTVLRAVVOHRLRCATNPYTEAN	136
QY	94	IDEVOKSDVSSTGGGVIDKDALGPMMLTALDGFEEVNL-EGNVESENVQYLRNOE	152
Db	137	YKPFLESD-----DELKHLILHAADGFLFVVGCDRGKILFVSESVKILNYSQN	185
QY	153	ELMAKSYVSIILHVDHFEVKNLLPKS-----IVNGSWSGEPPR	192
Db	186	DLIGOSLFEDYLHPMDIAK-VKEOLSSDPTAPRERLIDAKTGLPVKTDITPGSRILCSGAR	244
QY	193	RNSHTFCRMVLV-PLPDSSEEGHNDQNAHQEYEMOCFAVSQPRStK-----	239
Db	245	R---SFCRCMKCNRPVSKVEDKDFPSTCSKKADKSFCTIHTSTGYLISWPEPKMGLED	301
QY	240	---EEGEDLOSLICVARR---VPMK-ERPVLPSSSEFTTRDLOCKITSLD-TSTM	289
Db	302	NEPDNCGNL-SCLVAILGRILSHVYPQVYNEIRKMSHEVYSRHAIDKFPVVDRAAI	360
QY	290	AAAKPGMEDLYRRCIQKFHAGHEGESSYAKRHHHEVL--ROGLAESOIYRPSLSDGLV	347
Db	361	LAVLP-0ELLTSTCYEYFHO---DDIGHLAECHROYLQTREREKTL-TNCSYFKIKDGSFI	414
QY	348	AAQKSKLINSQTNENPOLVLSLMLHREQVWCVPNPLDTGQTMKRLPLISSNSPAAQA	407
Db	415	TLSRMSFEMNPWTKVEYIYS-----TIVVILANLEG---GDPFPOLTAAP-----	460
QY	408	LCSGNPGQDWTLISSNINFPINGPKEDOMPMRGSGSGGMHVSQ---MOATPQGSNYA	464
Db	461	-----HGMDSMLPGEGEGPKTHPTVLPDGTGTAAGACKIGRMIAEELMEIHRINGS---	512
QY	465	LKANSPOQSEFGANPGOFTSMLSPRHMSPG-----VAGSPRIIPSOFSWAGSLHSPVG	519
Db	513	-----SPSSCGSSPLNITS--TPPDASSPGKRIILNCGTDPIDISSGLSGAOENP---	562
QY	520	CSSTGNSHSTYNSLNLQALSEGHVSL-----GSSLASPD	566
Db	563	-----GYPSDSS-----SILGENPHGIDIMDQDSSSPSD	596

RESULT 25  
 US-08-816-693A-51  
 ; Sequence 51, Application US/0881693A  
 ; Patent No. 5874241  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Takahashi, Joseph S  
 ; APPLICANT: Turek, Fred W  
 ; APPLICANT: Pinto, Lawrence H  
 ; TITLE OF INVENTION: Clock Gene and Gene Product  
 ; NUMBER OF SEQUENCES: 53  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Dressler, Rocky, Milnow & Katz  
 ; STREET: Two Prudential Plaza, Suite 4700  
 ; CITY: Chicago  
 ; STATE: Illinois

```

1 COUNTRY: USA
2
3 ZIP: 60601
4
5 COMPUTER READABLE FORM:
6
7 MEDIUM TYPE: Floppy disk
8
9 COMPUTER: IBM PC compatible
10
11 OPERATING SYSTEM: PC-DOS/MS-DOS
12
13 SOFTWARE: PatentIn Release #1.0, Version #1.30
14
15 CURRENT APPLICATION DATA:
16
17 APPLICATION NUMBER: US/08/816,693A
18
19 FILING DATE:
20
21 CLASSIFICATION: 435
22
23 ATTORNEY/AGENT INFORMATION:
24
25 NAME: No. 5874241thrup, Thomas E
26
27 REGISTRATION NUMBER: 33,268
28
29 TELECOMMUNICATION INFORMATION:
30
31 TELEPHONE: 312-616-5400
32
33 TELEFAX: 312-616-5460
34
35 INFORMATION FOR SEQ ID NO: 51:
36
37 SEQUENCE CHARACTERISTICS:
38
39 LENGTH: 747 amino acids
40
41 TYPE: amino acid
42
43 STRANDEDNESS: single
44
45 TOPOLOGY: linear
46
47 MOLECULE TYPE: peptide
48
49 US-08-816-693A-51

```

Query Match	3.58;	Score 267;	DB 2;	Length 747;
Best Local Similarity	20.98;	Pred. NO. 6.7e-11;		
Matches 201;	Conservative 133;	Mismatches 345;	Indels 282;	Gaps 44;

QY	4	MGENTSDSRAATRRKRCBDQJGSPKNTKRRNHDENKTYIEELAEIIFANFNIDINF	63
Db	1	MDEDEKRAKRASRNK-----SEKKRRDGFNVLIKELTSLSPGNIR-----	41
QY	64	NFKPRKCALIKETVQIQIQIKQEKKAAANIDENQKSDVYSTGGGVYIDKDALGRMILEY	123
Db	42	--KMKITVLEKVISFLO--KNEYSAOTETCIDIDQDWKPS---FLSNEEFTQMLEAL	93
QY	124	DGFEEVNLDEGNVNVSENVTOYLRYNOBELKNKSYSILAHGDHTEFNKLIPKSYNG	183
Db	94	DG-FLAVTTDGSIIYVSDSIFPLHLHPSDVADQNLNLPRLQREHSEYVKKILSHMLYD	152
QY	184	GSWSEPPRRNSHTNCRMLYKRLPDSEBEGD-----NOEAKOKYETMOC---FAVS	233
Db	153	SPSPETLYKSDNDLEFYCHLLRSGLSMPKEPPEYTIKFVGNFRSYNVNVPSPSCNGEDNTLS	212
QY	234	QPKSIEKEGEDLOSLICAVARVP---KKERVLPSSSEFTTRDQJGITSIDTSTMA	290
Db	213	RPCRPLG-----KVCFITVRLATQFLKEMCVDERLEEFTRSHSLMKFLLD---HRA	265
QY	291	AMKPGM---EDLVRCIOKFNQAHGEGSVYAKRHHHEVLRQGLAFSQIYRFSLSIDGTLY	347
Db	266	PRIGLYLFEVLTGTSGYUYNL---DDELLARQHNMQGGKKSOCYRLLTGGQWI	321
QY	348	AAQTKSKILRSOTNEPQVLSLH-----MLHREDNVCYMNDLQGTQMGKPLNPIS	399
Db	322	WLQTYIYLYTHQWMSKPEFYCTHGSVSVADYRVRRQELAEDEP-----	366
QY	400	SNSPHQALGSGNPQODMTLSSNINPRLNGPEQGMPIRGGRGGSGMNVHSGMQATTPQ	459
Db	367	--PEAHSA-----KKDSLE-----PRO-----FNALDGASGLSP-	394
QY	460	GSNVALKMNSPQSSPGMNPQPTSMLSPRHRMSBVGASPRIP---PSQFSPAGSLHS	515
Db	395	-----SPSASRSSHKSHTAMSEPIIPTKLMESTALPRATLPQEL-----	437
QY	516	PVGWSSSTGNHSHYTNSSLNLQALSEBHGVLGSSLSLPDKMGNLQNSPVNNMPPIS	575
Db	438	PVG-----LSQATMPMLSSSC---DLTQQLDAPQTQLQSPAPQFS	473
QY	576	KMGSLSDKDCGLYGPESEGTGQAESSCHPEQKE-----TNDPNL-----PPA	620





Query Match 3.5%; Score 267; DB 4; Length 747;  
 Best Local Similarity 20.9%; Pred. No. 6,7e-11;  
 Matches 201; Conservative 133; Mismatches 345; Indels 287; Gaps 44;

```

OY 4 MGENTSDPSRAETRRKRECPDQGPSKRNTERKREOEKNKYEEELAEIFANFNDINF 63
  | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
  1 MEDEKDRAKRASRNK-----SEKRRDQNVLIKELSMPLPGIR----- 41
OY 64 NEKPKCALIKETVKOIROIKEOEKAAANIDEVOKSDVSTGQGVYDKDALGPMLEAL 123
  | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
  42 --KMKITVLEKVIQFLQ--KHNEVSAQTEICDIOQDMKPS---FLSNEEFTQMLEAL 93
OY 124 DGEFFVNLGNVNFVSENVTOYLRYNOEELMKSVSYSLHNGDHEFFKNLPRKIVNG 183
  | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
  94 DG-FIAVTTDGSITIVSDSITPLGLHPSDVMQNLNLFPEOEHSEVYKILSSHMLVTD 152
OY 184 GSWSGEPFRNRNHTFNCRLVPLPDSEEGHD-----NOEAHQEYTMQC---FAVS 233
  | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
  153 SPSPEYLKSDNLEFCHILRGSLNKEPTEYIKFVGNFRSYNNVPSPCGPNITLS 212
OY 234 QPKSIKEGEDLOSLICVARRVP---MKRPVLPSSSEFTTRDIOGKITSLDITSTMA 290
  | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
  213 RRCRVPILG---KVCFIATVRLATPOFLKMCVDEPLEEFTSRHSLEMKFLFD--HRA 265
OY 291 AMKPGM---EDLVRCIOKEFHNOHEGESVYAKRRHHHEVLROGLAFSQTIRPSLSDTIV 347
  | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
  266 PRTITLPEEVLGTSGYDYHT---DLELLARQHILMOPGKSCCYRFLTKGQOMI 321
OY 348 MAOTSKLRSQTNPEPOLVISLH-----MLHREOVNVCVNPNDLQTMGKPLNPIS 399
  | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
  322 WQTHYIYTHQMNKSPETIVCTHSVYADYRVERROLAEDP----- 366
OY 400 SNSPRHQALCSGNPCQDMTLSSNINPPIKPKQKMGKPRGGSGGNHNVSGMQATTPQ 459
  | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
  367 --PEHSA-----KKDSLE-----PRO-----FALGASGLSP- 394
OY 460 GSNVALKMNPSPOSSPGMNPQOPTSMLSPRHMSPGVAGSPRIP---PSOPSPAGSLMS 515
  | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
  395 -----SPASSRSRSHKSSHTAMSEPISTPTKLMASTALPRATLPDEL----- 437
OY 516 PVGVCSSGTGNSSHSTYNNSLNALQALSEGHSVLSGLSSLPADLMKGNLQNSPYVMNPPLS 575
  | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
  438 PVG-----LSQATMPLSSSC---DLTQQLQOTQIOLSPAPQPS 473
OY 576 KKGSLDKDCFLYGPESGTTGQAESSCHGEQKE-----TNDPML-----PPA 620
  | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
  474 AOFSM-----FOTIKDQLEORTIRILQANIR--WQOELHKTQEQOLCLVODSNVOMFLQOP 527
OY 621 V-----SSERADGOSRLHOSKQOTKILLQTLTSSDOMERPPLASSLSLDTKKDSTGSLPGSG 676
  | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
  528 VLSFSSSIORPAQOOLQORAAQPOLVO-----LQGOI 559
OY 677 STHTGSLKEKHILRLHLODSSSPVDLAKLTAEATGKDLSSOESSSTAPGSEVTIKQEPVS 736
  | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
  560 ST-----QVTHLHRESSV---ISGPRKMSQSLSGSSSLSPSSST----- 599
OY 737 PKKKENALIRYLDDKDTKIDGLPEI--TPKLERLDSKTDPAS--NTKLIAKTEKEENS 792
  | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
  600 -----LPRLTTPASTPDQSQCPSPDFHRLQLRLLSLQPIRP 637
OY 793 FERPD-----QPSSELNLEELIIDDQNSQLPOLFPDTRGAGASVDKAIINDMLQTLA 848
  | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
  638 MMPSGCDARQP--SEVSRFGQVYKAO--SQFPD--HPNSSVLLMG---QAVLHP--SFPA 688
OY 849 ENSPTPVGAOK-----TALRISQ-----STFNNRPGOLGRLLPNOLPLDITL--OS 895
  | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
  689 SPSPIQPMQAQOQPPRQAPATSLHSEODLSLSTFSQ--QGTIGLQOP--QPRRRVSLSS 746
OY 896 P 896
  | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
  747 P 747
  
```

RESULT 28  
 US-08-885-291-52  
 ; Sequence 52, Application US/08885291A  
 ; Patent No. 6057125  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Takahashi, Joseph S.  
 ; APPLICANT: Turek, Fred W.  
 ; APPLICANT: Pinto, Lawrence H.  
 ; TITLE OF INVENTION: CLOCK GENE AND GENE PRODUCT  
 ; FILE REFERENCE: 0290-5  
 ; CURRENT APPLICATION NUMBER: US/08/885,291A  
 ; EARLIER FILING DATE: 1997-06-30  
 ; EARLIER APPLICATION NUMBER: 08/816,693  
 ; NUMBER OF SEQ ID NOS: 55  
 ; SOFTWARE: Patentln Ver. 2.0  
 ; SEQ ID NO 52  
 ; LENGTH: 824  
 ; TYPE: PRT  
 ; ORGANISM: Mus musculus  
 ; US-08-885-291-52

Query Match 3.5%; Score 265; DB 3; Length 824;  
 Best Local Similarity 20.7%; Pred. No. 1.1e-10;  
 Matches 203; Conservative 143; Mismatches 391; Indels 244; Gaps 43;

```

OY 4 MGENTSDPSRAETRRKRECPDQGPSKRNTERKREOEKNKYEEELAEIFANFNDINF 63
  | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
  1 MEDEKDRAKRASRNK-----SEKRRDQNVLIKELSMPLPGIR----- 41
OY 64 NEKPKCALIKETVKOIROIKEOEKAAANIDEVOKSDVSTGQGVYDKDALGPMLEAL 123
  | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
  42 --KMKITVLEKVIQFLQ--KHNEVSAQTEICDIOQDMKPS---FLSNEEFTQMLEAL 93
OY 124 DGEFFVNLGNVNFVSENVTOYLRYNOEELMKSVSYSLHNGDHEFFKNLPRKIVNG 183
  | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
  94 DGFITAVTTDGSITIVSDSITPLGLHPSDVMQNLNLFPEOEHSEVYKILSSHMLVTD 153
OY 184 GSWSGEPFRNRNHT---FNCRLVPLPDSEEGHD-----NOEAHQEYTMQC---F 230
  | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
  154 -----SPSEYELKSDSDELEFCHILRGSLNKEPTEYIKFVGNFRSYNNVPSPCGPN 210
OY 231 AVSQSIKEGEDLOSLICVARRVP---MKRPVLPSS--ESFTTRDIOGKITSLDITS 286
  | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
  211 TLSRPCRV--PLKEV--CFIATVRLATPOFLKMCIVDEPLEEFTSRHSLEMKFLFD-- 265
OY 287 TMRAMKPGM---EDLVRCIOKEFHNOHEGESVYAKRRHHHEVLROGLAFSQTIRPSLS 343
  | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
  266 --HRAPRTIGYLPPEVLGTSGYDYHT---DLELLARQHILMOPGKSCCYRFLTKG 320
OY 344 GTLVAOTSKLRSQTNPEPOLVISLH-----MLHREOVNVCVNPNDLNG-----O 389
  | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
  321 QQWIMIQTHYIYTHQMNKSPETIVCTHSVYADYRVERROLAEDPSEALHSAK 380
OY 390 TWGKPLNP-----ISSNSPAHQALCSGNPCQDMTLSSNINPPIKPKQKMG 435
  | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
  381 DKSSLEPRQHNALDVGASGLNTHSPASSR--SSHKSSHTAMSEPISTPTKLMAEAST 439
OY 436 MEMGRFGSGGANNHVSGMATTPQ-----GSNVALKMNPSPOSSPGMNPQOPTSMLSPH 490
  | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
  440 PALPR-----SATLPQELPVGLSOAATMPAPLPS-----PLSCDLTQ 478
OY 491 RMSPPVAGSPRIIPPSQSPAGSLHSPVGVCSSTGNSSHSTNSLNALOA-----LSEHG 545
  | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
  479 LRPQYVLOSTPAPMAQFSQFSMFOYI-----KDLBQRTIRLQANINMOEELHK 529
OY 546 VSLGSSLASPDLKMGNIQNSPYNM--NPPPLSKMGLSKDCFLYGPESGTTGQAESS 603
  | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
  530 IOEQLCL-----VDSNVOMFLQOPAVSL-----SFSSIORPQAQOOLQORSA 572
  
```

QY 604 CHPEQKETNDPMLPPAVSSERADGSRLDHSGQTKLQLLTTRKSDOMEPSPLASS-LS 662  
 Db 573 AVTQPOLGAG-POLPGQISSAQTYSQHILRES-----SVISYQG----PKPMRSSQLM 620  
 QY 663 DTKNDSTGSL--PQSGTHTGSLKEKKHILHRLDSSSPVDLAKLAEATGKDLSESS 720  
 Db 621 QSSGRSSSSLVSPSSAT-----AALPSSLTLTPASTSODASQ--- 659  
 QY 721 STAGSEVTIKQEPVSPKKENALLRYLDKDDTKDIDGPEITPKLERLSDKTPA--SN 778  
 Db 660 -----CQP-SDFESHDRLLLSQ-----PIQPMGSCDARQPSSEVR 698  
 QY 779 TKLIAMTEREKEMSEFEGDQPGSELNLEETLDDLQNSQLPQI-----PDDTRPG 828  
 Db 699 TGRQVKYASQOTVFQNDAPRANS-----SSAPMPVLLMGQAVLHSPSPASQPS 747  
 QY 829 APAGSVKQALINDMLTAEVSPVTVGAOKTALRISQSEFNPRGQGRLLPNCNLP 888  
 Db 748 PLQPAQARQOPPOHYLVQA---PTSLHSEQDSSL- -STYSO-QPGTLAGYPPQAP 801  
 QY 889 LDITLQSPGTGAGPPPIRNS 909  
 Db 802 QPLR-----PPRRVSS 812

RESULT 29  
 US-09-496-672-52  
 ; Sequence 52, Application US/09496672  
 ; Patent No. 6291429  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Takahashi, Joseph S.  
 ; APPLICANT: Turek, Fred W.  
 ; APPLICANT: Pinto, Lawrence H.  
 ; TITLE OF INVENTION: CLOCK GENE AND GENE PRODUCT  
 ; FILE REFERENCE: 0290-5  
 ; CURRENT APPLICATION NUMBER: US/09/496,672  
 ; CURRENT FILING DATE: 2000-02-03  
 ; PRIOR APPLICATION NUMBER: 08/885,291  
 ; PRIOR FILING DATE: 1997-06-30  
 ; PRIOR APPLICATION NUMBER: 08/816,693  
 ; PRIOR FILING DATE: 1997-03-13  
 ; NUMBER OF SEQ. ID NOS: 55  
 ; SOFTWARE: Patentln Ver. 2.0  
 ; SEQ ID NO 52  
 ; LENGTH: 824  
 ; TYPE: PRT  
 ; ORGANISM: Mus musculus  
 ; US-09-496-672-52

Query Match 3.5%; Score 265; DB 4; Length 824;  
 Best Local Similarity 20.7%; Pred. No. 1.1e-10;  
 Matches 203; Conservative 143; Mismatches 391; Indels 244; Gaps 43;  
 QY 4 MGEITSPSRARETRKKRCPDOLPSKRNTEKRNREOKYIELLAELITANNDINDNF 63  
 Db 1 MDEDEKRAKASHNK-----SEKKRQDFNLILKLSMLGPNR----- 41  
 QY 64 NFKDKCAILKETVQKROIKEEKAANAANDEYKSDSVSTGGVGDIDKDLGPMLEAL 123  
 Db 42 --KDKKTIVLEKVIQFLD--KHNEVSAQTEICDIOODKPS-----FLSNSEFTQMLEAL 93  
 QY 124 DGEFFVNLGNNVVENVTQYLRYNOELMKNKSVYSLHVGDTHEFVNKLPRKSTYNG 183  
 Db 94 DGFIAVTGSGSIYVSDSTIRPLGHLPSDVMDQNLNLFPEQEHSEVYKILSSHMLVTD 153  
 QY 184 GSWGGEPPRRNSRT---FNCMLYKPLPDSEEGHD-----NDAHOKYETMQC----F 230  
 Db 154 ---SPSEYLSKSDSLEFYCHLLNGSLNPKKEFTYEVYIKFVGNRNNVNSPSCNGFDN 210  
 QY 231 AVSQPKSITKEGEDLQSCLCIVARRVP---MKERPVLS--ESFTTRDLOGKTTISLDT 286  
 Db 211 TLSRPRCV-PLGKEV--CFIATVRLATPQFLKMCIVDEPLEEFTSRHSLEWKFLLD-- 265

QY 287 TMRAMKGV---EDLYRRCIOKFHAQHESESYAKRHHNEVLROGLASFQIYRESLSD 343  
 Db 266 -HRAPIIGYLPFEVLGTSYDYHI-----DDELLARCHQHLMQFGIGSCCYRFLTKG 320  
 QY 344 GTLYAQTCKKLIRSGTTNEPOLYISLH-----MLHEQNVCVNPPDLTG-----Q 369  
 Db 321 QOWIMLOTHTYIYHQNMSPEFVCTHVSADVAVVERROELALDEPSEALHSALK 380  
 QY 390 TMGKPLNP-----ISSNPAHQALCSGNPGQDMTSSNINPINOPEQMG 435  
 Db 381 DKGSSLEPRQHFNALDVGAGLNTSHPSASR--SSKSSHTAMSEPTISPTKLMAEAT 439  
 QY 436 MPMGFRGSGGMNHVSGMAATTPQ-----GSNALKNNSPQSSPGNPGQPTSMSPRH 490  
 Db 440 PALPR-----SATLPQELVPGISQATAPALPS-----PLSCDLTQ 478  
 QY 491 RMSGVAGSRIPRPSQSPAGSLHSPVGVSSSTGNSHSTYNSLSMLAQ-----LSGCHG 545  
 Db 479 LLPQTVLQSTPPAPMAQPSAQSMTFTI-----KDLEORTRLIQANIRMQOEELK 529  
 QY 546 VLSGSSLASPDLMKGNLQNSPVNM--NPPLSKMSLSDKDCFLYGESEGTGQAES 603  
 Db 530 IQEQLCL-----VQDSNVQMFLOQPAVSL-----SPSSIORPAPQOOLQGRSA 572  
 QY 604 CHPEQKETNDPMLPPAVSSERADGSRLDHSGQTKLQLLTTRKSDOMEPSPLASS-LS 662  
 Db 573 AVTQPOLGAG-POLPGQISSAQTYSQHILRES-----SVISYQG----PKPMRSSQLM 620  
 QY 663 DTKNDSTGSL--PQSGTHTGSLKEKKHILHRLDSSSPVDLAKLAEATGKDLSESS 720  
 Db 621 QSSGRSSSSLVSPSSAT-----AALPSSLTLTPASTSODASQ--- 659  
 QY 721 STAGSEVTIKQEPVSPKKENALLRYLDKDDTKDIDGPEITPKLERLSDKTPA--SN 778  
 Db 660 -----CQP-SDFESHDRLLLSQ-----PIQPMGSCDARQPSSEVR 698  
 QY 779 TKLIAMTEREKEMSEFEGDQPGSELNLEETLDDLQNSQLPQI-----PDDTRPG 828  
 Db 699 TGRQVKYASQOTVFQNDAPRANS-----SSAPMPVLLMGQAVLHSPSPASQPS 747  
 QY 829 APAGSVKQALINDMLTAEVSPVTVGAOKTALRISQSEFNPRGQGRLLPNCNLP 888  
 Db 748 PLQPAQARQOPPOHYLVQA---PTSLHSEQDSSL- -STYSO-QPGTLAGYPPQAP 801  
 QY 889 LDITLQSPGTGAGPPPIRNS 909  
 Db 802 QPLR-----PPRRVSS 812

RESULT 30  
 US-08-816-693A-52  
 ; Sequence 52, Application US/08816693A  
 ; Patent No. 5874241  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Takahashi, Joseph S.  
 ; APPLICANT: Turek, Fred W.  
 ; APPLICANT: Pinto, Lawrence H.  
 ; TITLE OF INVENTION: CLOCK Gene and Gene Product  
 ; NUMBER OF SEQUENCES: 53  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Dressler, Rocky, Milnamow & Katz  
 ; STREET: Two Prudential Plaza, Suite 4700  
 ; CITY: Chicago  
 ; STATE: Illinois  
 ; COUNTRY: USA  
 ; ZIP: 60601  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentln Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:

```

: APPLICATION NUMBER: US/08/816,693A
: FILING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: NO. 5874241thrup, Thomas E
: REGISTRATION NUMBER: 33,268
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 312-616-5400
: TELEFAX: 312-616-5460
: INFORMATION FOR SEQ ID NO: 52:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 824 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: US-08-816-693A-52

```

Query Match 3.4%; Score 262; DB 2; Length 824;

Best Local Similarity 20.7%; Pred. No. 1,8e-10; Matches 203; Conservative 142; Mismatches 392; Indels 244; Gaps 43;

```

QY 4 MGENTSDPSRAETRRKRECPDGLSPKRNTEKRNREQENKYEIELAFANFNIDNF 63
DB 1 MDEDEKRAKASRNK-----SEKKRRDQFNVLLIKELSMLGNNR----- 41
QY 64 NFKDKCAILKEVYKQIRQIEQEKAAANIDEVQKSDVSTGGGVIDKDALGPMLEAL 123
DB 42 --KMDKTVLEKVIQGLD--KHNEVSQTEICIDIOQDKPS---FLSNEFTOLMLEAL 93
QY 124 DGEFFVNLGEGNVFVSENTQYLRNQEELMANKSVSILHVGDTHEFVKMLLPRIYNG 183
DB 94 DGIATATTDGSIITVSDSTPLGLHLPBVDMDONLFLPEQERSEYKILLSSMLVTD 153
QY 184 GSWSGEPPRRNSH--FNCRLVYKPLPDSSEEGHD-----NOEHQKYEYTMOC---F 230
DB 154 ---SPSEYILKSDSDLEFYCHLNGSLNPKFEFTEYIKFVGNFRSYNNVSPSCNGFDN 210
QY 231 AVSPKRIKEGEDLOSLCIVARRV--MKERVLPSS--ESTTRDLOGKITSLDTS 286
DB 211 TLSRRCV--PLGKEV--CFATVRLATPQFLKMCIVDEPLEEFTSRSLMKFLFD-- 265
QY 287 TMRAMKPGW---EDLVRCIQKFAHQEGESVYAKRHHEVLRQGLAFQIYRFSUSD 343
DB 266 --HRAPITIGYLFPEVLGTSGVDYHT---DDELARCHQHLKQFGKSCCRFLTKG 320
QY 344 GTLVAAQTKSLIRSQTTNEPOLVLSLH-----MLHRDNYCVNPDLTG-----Q 389
DB 321 QOWIMLQTHYIYTHQNNKREPIVCFHSVSYADVVERERQELALEDPSPSEALHSSALK 380
QY 390 TMRKPLRP-----ISSNPAHQALCSGNGQMTISSNINFPINGKEQMG 435
DB 381 DKGSLEPRQHFNALDVGASGLNTHSPSASSR--SSHKSSTAMSEPISTPTKLMAEAST 439
QY 436 MPMRPFSGSGMNHVSGMOATTPQ-----GSNYALKMNSPQSSPGMNPQPTSLSPRH 490
DB 440 PALPR-----SATLROELVYRGLSQATMPAPLP-----PLSCDLTQ 478
QY 491 RMSPGVAGSPRIIPSPSPAGLSHPVGVCSGTNSHSTNSLNAQO-----LSECHG 545
DB 479 LIPQTVLQSPAPMAQFSQAFSMFQTI-----KDQLRQTRILLQANIRMQOELHK 529
QY 546 VSLGSSLASDPLKMGNLQNSPVNM--NPPLSKMGSLSKSCFGLGEPSPSGITGOASS 603
DB 530 IQEQLCL-----VQDSNVQMFLOQPAVSL-----SFSIORPEAQOOLQORSA 572
QY 604 CHPEQKRETNPNLPPAVSSERADQSRHLSKQTKLLQLITRKSDMEPSPLASS--LS 662
DB 573 AVTPQPLGAG--POLPQGLISSAQVTSQHLRES-----SVISTGQ---PKPARSSOLM 620
QY 663 DTNKDSFGSL--PGSGSTGTSLKEKKHILHRLLODSSSPVDLAKLTAEATGKDLQSSS 720

```

```

DB 621 QSSGRGSSSLVSPSSAT-----AALPSPNLTPPASTSQDASQ--- 659
QY 721 STAPSEVTIOEPVSPKKKENALLRYLDDKDTDIGLPETPKLERLDSKTDA---SN 778
DB 660 -----CQP--SPFSDROLRLLSQ-----PIQPMPPSCDARQPSSEVR 698
QY 779 TKLIAMKTEKESMSPEDQSGSELNLEILLDQNSQLPOL-----FPDTRPG 828
DB 699 TGRQVKYASQSTVPCPNPAHPANS-----SSAPMVLKMGAVLHPSFPAQPS 747
QY 829 APAGVDKQAIINDLMQLTAENSPVTPVGAQKTALRISQSTFNNRPQGLRLPNQMLP 888
DB 748 PLQPAQARQPPQHYLYQVA---PTLSHQODSILL--STYSQ--QPTLGYPPQPPAQP 801
QY 889 LDITQSPTAGPPPIPNSS 909
DB 802 QPLR-----PPRRVSS 812

```

RESULT 31

US-08-785-310A-7

```

: Sequence 7, Application US/08785310A
: Patent No. 5840532
: GENERAL INFORMATION:
: APPLICANT: McKnight, Steven L.
: TITLE OF INVENTION: Neuronal PAS Domain Protein
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESS: SCIENCE & TECHNOLOGY LAW GROUP
: STREET: 268 BUSH STREET, SUITE 3200
: CITY: SAN FRANCISCO
: STATE: CALIFORNIA
: COUNTRY: USA
: ZIP: 94104
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patent Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/785,310A
: FILING DATE: 21-JAN-1997
: CLASSIFICATION: 536
: ATTORNEY/AGENT INFORMATION:
: NAME: OSMAN, RICHARD A
: REGISTRATION NUMBER: 36,627
: REFERENCE/DOCKET NUMBER: UTSO:1226
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 343-4341
: TELEFAX: (415) 343-4342
: INFORMATION FOR SEQ ID NO: 7:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 824 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: US-08-785-310A-7

```

Query Match 3.4%; Score 259; DB 2; Length 824;

Best Local Similarity 19.4%; Pred. No. 3e-10; Matches 215; Conservative 154; Mismatches 399; Indels 338; Gaps 46;

```

QY 4 MGENTSDPSRAETRRKRECPDGLSPKRNTEKRNREQENKYEIELAFANFNIDNF 63
DB 1 MDEDEKRAKASRNK-----SEKKRRDQFNVLLIKELSMLGNNR----- 41
QY 64 NFKDKCAILKEVYKQIRQIEQEKAAANIDEVQKSDVSTGGGVIDKDALGPMLEAL 123
DB 42 --KMDKTVLEKVIQGLD--KHNEVSQTEICIDIOQDKPS---FLSNEFTOLMLEAL 93

```

```

QY 124 DGEFVNVNENGVENVSENVTOYLRNQEELANKSVYSIIHGDHTEFVKNLLPKSIVNG 183
  ||| | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 94 DGFIIAVTTDGSIIYSDSTITPLGHLPSVDMQNLNLFPEOHESEVYKILSSHMLVTD 153
QY 184 GMSGSGPPRRNSHT---FNCRMLVFKLPDSEEGHD-----NOEAKHOKETMOC-----F 230
  ||| | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 154 ---SPSEPIKSSDLEFCHLGRSINKEPPTYEIKFVGFRSTNNVSPSCGFNDN 210
QY 231 AVSQPSIKEEGEDLOSCLICVARRVP---MKERPVLPS--ESFTTODLOGKITSLDTS 286
  ||| | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 211 TLSRCPRV-PLGREV--CFIATVRLATPOLKCMCIYDELEFEFTSHSLKEMFLFLD-- 265
QY 287 TMRAMKPGW---EDLVRCIOKFNHOGESVYAKRHHEVLRGCLAFSQTIRPSLSD 343
  ||| | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 266 -HRAPPIIGLPEFVLTSGSYDYTHI---DDELLARCHOHLMOGTGKSCCYRFLTKG 320
QY 344 GTLVAQTSKILRSQTTNTPOLVLSHMLHREQNVCMNPDLTGOTMCKPLMPISNSP 403
  ||| | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 321 QOMIMIQTHYITYTHOMNSKPEFTVCTHVSYSTADVRV-----E 339
QY 404 AHOALCSGNFGQDMTLLSNINFPINGKEQMGPMGRFG---GSGGMNHFSGMOATTPQ 459
  ||| | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 360 RQELALEPDPSEALHSSALK-----DKSSLEPRQHFNALDVGASGLN----- 403
QY 460 GSNYALKMNSPSSQSPGMNPGOPTSMISPRHRMSP--GVAQSPRIIPSPSPKASLHSPY 517
  ||| | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 404 -----TSHSPSASSRSSSHKSSHTAMSEPTSTPTKLAEMASTPALPRSATLPQ---ELPV 454
QY 518 GVCSTGNSHSYTNSSLNALQALSEGCVLSGLSSLASP---DLKMGVLQSPVNMNPPPL 574
  ||| | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 455 -----PGLSQ--AATWAPLPSPSLCDLTQOLLPOTVIQTSTPAPM 492
QY 575 SKMGLSDSKDCFLGYGEPESEGTGAESSCHPEOKETNDPNLPVAVSSERAD---GQSR 631
  ||| | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 493 AGPSAQ-----FEMF-----QTIQDLEQRTIRLIQANIRMOQEE 526
QY 632 LHSKQGTKLQLLTTKSDMERSPLASSLSDTNKSDTSGSLPGSGTHGTSLEKHKILH 651
  ||| | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 527 LHKIQEOLCLVQDSNVOMFLQPA--VLSFSSSTQRPFA-----Q 564
QY 692 RLLQDSSSPVDLAKTLTAATAGTKDLSOESSSTAGSEVTIKOEPV---SPKKENALLRY 747
  ||| | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 565 QQLQQAASAVTQPOLGA---GPOLPGQISSAQVTSOHLRESEVYISQGRKPMRSSQLQ 621
QY 748 -----LLDKDDTKDIGPEITPKLERLDSKXTDPASNTKLAMKTEKEMSFEEDQ 799
  ||| | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 622 SSGRSSSIVSPSSATAALP---PSL---NLTPASTS-----QD 656
QY 800 GSELNDLEFLDLQNSQLPQLFPDTRPGAPASSVDKQALINDLMQLTANSFYTPVGAQ 859
  ||| | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 657 ASQCOPSPDSSHRODLRL--LSQPIOPMP--GSCDAR-----OPSEVSRTGRQ 702
QY 860 KTLARISOSTFNPRPGQLRLPLNQNLPLDITLQSPTAGPPPIRINSFPYVIPOQM 919
  ||| | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 703 VKTAQ--SQYTFQMP-----DAHPANSSSAPMYL----- 730
QY 920 MGNQMGINGNGLNCSGTMGINSASRPTMPSGEMAPQSSAVRYTCAATTSAANRPYQG 979
  ||| | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 731 -----LMGQAVLHPSFPASQ----- 745
QY 980 MIRNPASITMRPSSQDGOQOTLOSQVYMNIGPSELEKMNMGPOYSQO-----QAPPNQ 1032
  ||| | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 746 ---PSFLQPAQARQOPQPHY--LQVQAPTSLSHSEODSLSTYSQOPGTIGYPOPPAQ 800
QY 1033 TAPWPSIILPIDQAFASON---ROP 1055
  ||| | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 801 ----POPAPRRRVSSLSSESSGLQDP 822

```

RESULT 32  
 US-08-061-376-5  
 ; Sequence 5, Application US/08061376  
 ; Patent No. 6175000

```

; GENERAL INFORMATION:
; APPLICANT: Evans, Glen A.
; APPLICANT: Djabbali, Malek
; APPLICANT: Selleri, Lucia
; APPLICANT: Parry, Pauline
; TITLE OF INVENTION: CHARACTERIZATION OF A CHROMOSOME 11Q23
; TITLE OF INVENTION: TRANSLOCATION BREAKPOINT ASSOCIATED WITH ACUTE LEUKEMIAS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESS: Pretty, Schroeder, Brueggemann & Clark
; STREET: 444 South Flower Street, Suite 2000
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/061,376
; FILING DATE: 13-May-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Reiter, Stephen E.
; REGISTRATION NUMBER: 31,192
; REFERENCE/DOCKET NUMBER: P41 9387
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619)546-4737
; TELEFAX: (619)546-9392
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3969 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-061-376-5

```

```

Query Match 3 4%; Score 258.5; DB 4; Length 3969;
Best Local Similarity 18.6%; Pred.No.3.9e-09;
Matches 350; Conservative 243; Mismatches 611; Indels 673; Gaps 91;

QY 10 DPSRAETRRKKECPDOLGSPKRNTERKRNREQNKYIEELIAELIFANFNIDINFNEKPPK 69
  ||| | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1672 NPETESIPSRSSPE--GPPPVLTGVSKODDQPL-----DLGVKRRKMQ 1716
QY 70 ---CALIK--ETVKQIRQIKEDOKAANID---EVOKSDVSSTGQGYIDKDALGPMK 119
  ||| | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1717 GNYTSVLEFSDDIVKILQ-----AATNSDGOPEIKKAN-----SM 1752
QY 120 LEALDGEFFVNLGNVNVFSENVTOYLRNQEELANKSVYSIIHGDHTEFVKN--LLPK 178
  ||| | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1753 VKS---FTRQMERVFPWFSVYKSSRWEPNK-----VSSNGMLPNAVLPP 1795
QY 179 SI-VNGSGWSGEPRRNSHTFNCRLVYKPLPDSEEGHDNOEAHOKYETWQCFVAGQPS 237
  ||| | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1796 SLDHNTAQW--QRENSHTEQPPKAKITIPAKPKGPGRPDPTPLHPTPTILSTDS 1853
QY 238 IKKEG-----EDLOSCLICV-----ARRVPKERP-----VLPSSSEFTTR 273
  ||| | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1854 REDSPELNPPGIEDNRQCALCLTYDDDSANDAGRLIYIGQNEWTIVNICALMSAEV--- 1910
QY 274 QDLOGKITSLDTSTMA-----AMKPGWE--DLVRCIOKFNHQAHE-----GESV 316
  ||| | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1911 EDDDGSLKNVHMAVIRGKQLRCEFCOKPGATVGCCLTCSNTHFMCSTRAKNCVFLDKK 1970
QY 317 SYAKRHH-----HEVLRQ-----GLAFSQIYRF----- 339
  ||| | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1971 VTCQHRDLIKGEVVPENGEFEVFRVVDPEGISLRKRLNGLEPENIHMMIGSMITDCL 2030

```

```

QY 340 ----SLDGTIVAQTK-----SKLIRSQTFNEPOLVILSHLHREOVNCPANPDLT 387
D 2031 GILNDLSD-----CEDKLEFPIGYCSRYWSTTDARKRCVYTCIKIV--ECRPVVEPDI- 2082
QY 368 GQTMCKPLNPISNSPAHQALCSGNPGQDMTSSNINFPINGPKQMGKMPKRGCSGGM 447
D 2083 NSTVHEDNRTIAHSPFTFTESSSKESQNTA-----EIIISPPDBRPHSQTSGSCY 2135
QY 448 NHVSGM-QATPGQSNVYALKNKNSPSSQSGMNP-----GQPTSMLSRRH-----MSF 494
D 2136 HVISVVPKIRTPS-----YSPQKSPGCRPLPSAGSP--PRTHEIVTVGDPLSS 2184
QY 495 GV--AGSPRIPSPQSPAGS--LHSPVGVCSSTGNSHTYNS-----533
D 2185 GLRSTGSRHSTSLSPGRSKRIMSPW--RIGNTYSRRNNVSVSTGTATDLESSAK 2240
QY 534 -----LMAQLSEHGVSGLSSLASPLDKMGNLONSPVNNMPPPLSKGS--LDS 582
D 2241 VVDHYLGLPNSSTSL--GQNTSTSSNLQRTVTVGN-KNSHLDCSSSEMKOSSASDIYS 2297
QY 583 KDCFLGYE-----PSEGTQOAESSCHPGQKEINDNLPRAVSEERADGSRHLDSKG 637
D 2298 KSS-SLKECKTKVYSSKSSBSAHNAVPG-----IPKLAP-----2332
QY 638 QTKLIQLITTKSDQMEPSPLA-----SLSDTNKNDSTGLPGSGTGTSLKEKHILHR 692
D 2333 -----QVHNTSRELNVSKISFAEPSSVSFSKE-----ALSFLHLRG 2373
QY 693 LLQSSSPVDLAKLTAEYTGDLDSOESSVAPGSEVTTKQEPVSPKKENALLRYLLKD 752
D 2374 QNRNRDQHTD-----STQSANSSPDEDETEVKTUKLSGMSNRSSII-----2413
QY 753 DTKDGLPEITPKLERLDSKTPASNTKLIMKTEKEMS-----PEPDQCGSELDNLE- 807
D 2414 -NEHWG-----SSSDROROKGKCKEYFEKHSSKSELEPQVTTGEGNLKP 2461
QY 808 EILDLOLQSPOLPEPD-----TRPGAP-----AGSVD-----835
D 2462 EFMEV-----LTPPEVMGQPCNNVSSDKIGDKLSMPGPKAPPMQVESAELQAP 2514
QY 836 -KQALINDMOLTAEN-----SPVTPVGAQKATLKRISOSTFNNPRPQGLRLLP 883
D 2515 RKRRTVYKLTPLKMEENESQNALKESSPASPLQIESTSPTEPISASENPGDPAVQPS 2574
QY 884 N-----ONLPL-DITLQSPGAP-----PP-----P1R 906
D 2575 NNTSCQDSQSNNTQWLPQDNRNLMLPDGPKPOEDGSEFRKRRTPRRSARASNNFGLTPLY 2634
QY 907 NSSPY-----SVIPOPGMGNOGIGNOGLGNS-----STGMIG 941
D 2635 GVRSYGEEDIPFYSSSTGKKRKSABEQVVGADDLSTDEDDLYYFTRTVYISSGEE 2694
QY 942 NSASR-----PMPSGEMAPSSAVRVTCATTSAMNR-PVOCGMIRNPAASIP 989
D 2695 RLASHNLFREEOCDLPKISOLDVDGTESTDVTATTKRSSQIPKRNGK-ENGTEMLK 2753
QY 990 MRPSQPGQROTLQ-----SOVMNITGPSELNMGGPQYSQ-----1025
D 2754 IDREPDAGEKEHVTKSSVGHKNEPKMNCCHSVRYKTOGGQSLSSLESRRVHTST 2813
QY 1026 -----OQAPPMQTAWPPEISILPIDQASFASON--ROFGSSPDLILCPH 1067
D 2814 PSDKNLDTYNTLLKSSDSDNNNSDGCNILLPSDLMDFVLKNTPMQALGESP-----2866
QY 1068 PAASPSDEGALLDQLYALRNFDGLF-EIDRALGIPELVVSQSAVDEQFSSQDSNIML 1126
D 2867 -----ESSSE-----LNLLEGGGLDSDNREKMDGLEFVFSQQLPTTEPVSSVSSISA 2916
QY 1127 EQ-----KAPVFPQOYASQAQMAQGSYSMPQDNFTMQR-----P 1163
D 2917 EEOGELPELPSDLSVLTTRSPFVPSQNPSPRLAIVSDS-----GKKRVTTIEK 2964
QY 1164 SVATLRMPQ---RGLRPTGLVONQNLRLQLHRLQAOONROPIMNQI--SNVSNVNL 1218

```

```

D 2965 SVASSESDPALLSPGVDPDPREGHMTDPDH---IQGHMDADHITSPPCGSVEQGHGNNQDL 3021
QY 1219 TLRRGVP-TQAFINMOMLAQORELLNQ-----HLR---QOMHQ 1255
D 3022 TENSSTPGLQVPSPTVPIONOKYVNSTDSPSQISNAVAQTTPPHLKPAETKLIYVN 3081
QY 1256 QOVQO-----RTL-----MMRGGLNMTP-----SMVAPSGMPTMS--NFRIPA 1294
D 3082 QMOPLYVLQTLPLNGVTKIQLTSSVSTPSVMEETNTSVLGMGGGLTTLGLNLSPLPS 3141
QY 1295 N-----AOQFPEPPN-----YGISQOPDPCF-----1315
D 3142 QSLPFSASKGLLPMHQHLSFPRAQTSSFPFNISNPPSGLLIGVPPDPQVLVSSS 3201
QY 1316 -----TGATTPQSPPLMSPRMAHTOSPMMQO--SQANPAYQAPSDI-----1353
D 3202 QRTDLSTTVAPPSQGLKRRPISRLQTRKXKLAPSSFTSPSNIAPSDVSNMTLINFTPSOL 3261
QY 1354 -NGMAQNMNGSMFSPQSPHFGQAMTS-MYSNNMNIIVSMATNTGSMSSMNMQTOQI 1411
D 3262 PNHPSLIDLGSLNTSSHRTVPNITKRSKSIYFEPAPL--LPQSVGTATTAAGTSTI 3318
QY 1412 SM--TSVTSVSTSGLS 1426
D 3319 SODTSHLTSGSVSGLAS 3335

```

```

RESULT 33
US-08-194-468-2
; Sequence 2, Application US/08194468
; Patent No. 5750336
;
GENERAL INFORMATION:
APPLICANT: Montminy, Marc R.
TITLE OF INVENTION: ASSAYS FOR THE IDENTIFICATION OF
TITLE OF INVENTION: COMPOUNDS WHICH INHIBIT ACTIVATION OF CAMP AND MITOGEN
TITLE OF INVENTION: RESPONSIVE GENES
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
STREET: 444 South Flower Street, Suite 2000
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/194,468
FILING DATE: 10-FEB-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Reiter, Stephen E.
REGISTRATION NUMBER: 31,192
REFERENCE/DOCKET NUMBER: P41 9672
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619)-546-4737
TELEFAX: (619)-546-9392
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2441 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-194-468-2

```

Query Match 3.3%; Score 252; DB 1; Length 2441;  
Best Local Similarity 20.6%; Pred. No. 5.3e-09;  
Matches 248; Conservative 142; Mismatches 442; Indels 372; Gaps 61;









```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/235,217
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/480,473
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile, Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 07265/053001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 805 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-235-217-4

```

```

Query Match          3.2%; Score 248; DB 4; Length 805;
Best Local Similarity 19.3%; Pred. No. 1.8e-09;
Matches 178; Conservative 144; Mismatches 322; Indels 280; Gaps 39;

```

```

Qy 39 REDEKTYIEELAELEFANFNDIDNENFKPDKCALIKETVYKOIROKEDEKAAANIDEVQ 98
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 9 RSKSEVEFELEAHOPLPHNVSSH-----DKASVRLTISTYL-VRKLLDAGDLDIEDDM 63
Qy 99 KSDVSTGOGYIDKDALGPMLEALDGFVFNLEGNVFEVSENYTOYLRYNOEELAMKS 158
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 64 KAQNMNC-----FYKALDGFVNYLTDDGDMIIYSDNNKMKGLQFELTGH 110
Qy 159 VYSILHVDHTEFVKNLPLKSTLVNGSGSGEPFRNSHTFNCRLMVKPLPSEEGHDNQ 218
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 111 VFDTTHPCDHDEMRLTHR---NG-----LVK-----KGKRON 141
Qy 219 EAHQYETMOCEFAVSQPSIKEEG-----EDLQSLIC 251
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 142 TORSEFLRMKCTLTSGRTMNIKSAWKVYLHCTGHIHYDTNNSNOPCGYKKPMTCLVL 201
Qy 252 VARRVPMKEREVLV-PSESEFTTRODLOGKITSLDTS-TMRAAMKPGMEDLVRCIOKPHA 309
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 202 ICEPIPHNSNTEIPLDSTFLSRHSLDMKFSYCDERITELMGEYF-ELLGRSLEYEY- 258
Qy 310 QHEBESVYAKRHHHEVLROGLAFSQIYRFSLSDGLVAQAOTKSKLIRSQTNEPOLVIS 369
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 259 -HALDS-DHLTKTHDMFTKGOVTTGQYRMLAKRGYVWVEQATVIVYNTKNSQOCIVC 316
Qy 370 LH-----MLHRQNCVNNP-DLTGOTMCKPLNPLSSNSPAH-----QAL 408
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 317 VNYVVGIIIOHDLIFSLQOTECVLKPVSSDMKMTOLFTKVESEDTSLFLKLEKPEAL 376
Qy 409 CSGNPGODMTLSNINPFIINGPKQMGMPGRFGSGGMNHYSGMQATTPOGSNYALKMN 468
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 377 TLAPAGDTIIS-LDF-----GSNDTEIDD 401
Qy 469 SPSSGSSPGMNGOPTSMKSPRRHM-SPGVASPRIPSPQSPAGSLHSPGVGCSSTGNSH 527
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 402 QOLEVEVPLYN---DVMLSPSPNEKIONINIAMSP-----LPATETPRPL-----R 442
Qy 528 SYTNSSTNALDALS-EGGVSGLGSSLASPDLMKGNLONSPPVNMNPPRLSKMGSLDSKDCF 586
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 443 SSADPALNOEVALKLEPSELESLETFMPOIO-----DQTP----- 478

```

```

Qy 587 GLYGPSEGGTTGOAESCHPGEOKE-----TNDENLPVAVSSERADGQSRHLDKSG 637
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 479 ---SPSGGSTRQ---SSPEPNPSSEXCFCYVDSDMNMEKLELVLEFLAEDTEAKKPFSTQ 532
Qy 638 OTKL-LQLIT-----TKSDMEPSPLASSLSDPTNKDSTGSL-----PG 674
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 533 DTDLDEMLAPYIPMDDDFQLRSFDQL--SPLESSASPEASPSQSTVTFVQQTQIQPT 590
Qy 675 SSGTHGTSIAKEKHKLH-----RLDSSSPVLAULTAEAT----- 711
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 591 ANATTTTATTDLTKVTYKDRMEDIKILIASPSPIHKEETYSATSPYRDTQSRASPFR 650
Qy 712 -GKDSQESSSTAPGS-----EYTIKQEPVSPK--KENALLRYLLDKDPT---KD 756
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 651 AKGVIEQTEKSHRSPNVLSVALSQRTTYPEELNPKILALQNNQRKMHDSGLFOA 710
Qy 757 IGLPEITPKLERLDSKTDTPASNTKLIAMKTEKEMSPFGDQPGSELNLEIIDL--- 813
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 711 VGIGTL-----LQOPDDHAATTSL---SWKRVGCKSSSEQNGMEQKTIILIPSDLACR 760
Qy 814 -----QNSOLPLFP-DTRPGAP 830
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 761 LIGQSMDESLPQLTSTYDCEVNAIP 784

```

```

RESULT 38
PCT-US96-10251-4
; Sequence 4, Application PC/TUS9610251
; GENERAL INFORMATION:
; APPLICANT: The Johns Hopkins University School of Medicine
; TITLE OF INVENTION: HYPOXIA INDUCIBLE FACTOR-1 AND METHOD OF USE
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/10251
; FILING DATE: 06-JUN-1996
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile, Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 07265/053001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 805 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US96-10251-4

```

```

Query Match          3.2%; Score 248; DB 5; Length 805;
Best Local Similarity 19.3%; Pred. No. 1.8e-09;
Matches 178; Conservative 144; Mismatches 322; Indels 280; Gaps 39;
Qy 39 REDEKTYIEELAELEFANFNDIDNENFKPDKCALIKETVYKOIROKEDEKAAANIDEVQ 98
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 9 RSKSEVEFELEAHOPLPHNVSSH-----DKASVRLTISTYL-VRKLLDAGDLDIEDDM 63

```



```

Db 478 MPOIDDP---ASPS-DGSTR---QSSP--EPNSPSEXC-----FDVSDMVNFK-- 519
Qy 541 SEGHSVLSGSSIASPDLKMGN-LQNSPVNMNPPLSKMGSLSDKDFGLYGEPSGCTGO 599
Db 520 -----LLEVERKLFADTEKKNPFSTQDIDLEMLAPYIPMD--DDFOL----- 561
Qy 600 AESSCHPEQKETNDPNLPPAVSSERADQSRILHDSKGFLLQLLTTRKSDMEPSPLAS 659
Db 562 --RSPDQLSPLESNSPS--PPMSVTGTGFOQTOL-----QKPTITAT 599
Qy 660 SLSDPNKSTGSLPSGSGHGTSLKEKHILRLLODSSPYDLAKLTAEAT----- 711
Db 600 ATTATTTES-----KTEYKDNKEIKIL--IASPSTQVPOETTAKAAYSQTHSR 650
Qy 712 -----GKDLSESSSTAPGS-----EVTIKOEVSFK--KRENALLRYLLDKD 753
Db 651 TASPBACKRVTEQTDKAPRSLNSATLNONRYPEELNKNTIASOAKKRRKMHG 710
Qy 754 T--KDGLPEITPKLERLDSKTDSPASNKLIAKTEKEMSPEDDQPSGLDNLEETLD 811
Db 711 SLFQAAGIGTL-----LQDPGDCAPTMSL--SMKRVKGFISSEONGTEGKTIILIPS 760
Qy 812 DLQNSQLQLPEDTPRGAPAGSVKQAIINDLMOLTAENSPV--TVGGAOKTLR 864
Db 761 DLACRLGQ-----SMD-----VSGLPOLTSYDCEVNAPIQSGSRNLLQ 798

```

## RESULT 40

```

US-08-045-806-4
: Sequence 4, Application US/08045806
: Patent No. 5378822
: GENERAL INFORMATION:
: APPLICANT: Bradfield, Christopher Alan
: APPLICANT: Dolwick, Kristin Marie
: APPLICANT: Poland, Alan
: TITLE OF INVENTION: An Receptor cDNA and Method of
: TITLE OF INVENTION: Determining Human Risks To Environmental Pollutants
: NUMBER OF SEQUENCES: 23
: CORRESPONDENCE ADDRESS:
: ADDRESS: Tilton, Fallon, Lungmus & Chestnut
: STREET: 100 South Wacker Drive, Suite 960
: CITY: Chicago
: STATE: Illinois
: COUNTRY: USA
: ZIP: 60606-4002
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/045, 806
: FILING DATE: 19930408
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Feitress, Susan B.
: REGISTRATION NUMBER: 31,327
: REFERENCE/DOCKET NUMBER: NU-9207
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (312)-456-7776
: TELEFAX: (312)-456-7776
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 848 amino acids
: TYPE: AMINO ACID
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-045-806-4

```

```

Query Match 3.1%; Score 235; DB 1; Length 848;
Best Local Similarity 18.2%; Pred. No. 1.7e-08;
Matches 188; Conservative 156; Mismatches 371; Indels 320; Gaps 42;

```

```

Qy 4 MGEINTSDPSRAETRRKRECPDQAGSP-----KRNTEKRNREQENKYELELAI--FANFN 58
Db 1 MNSSSANITTYAARRKRRKPVQKTVKPIPAEGISNSKRRDRLNLELDLALFLFP--Q 58
Qy 59 DIDNENFKPKDCAILKETVQKIQKIEOEKAAANIDEVOKSDVSTGGCVID-----K 112
Db 59 DVIN--KDKLSLYRLSVSYLR-----AKSFDVALKSSPIERNQONCRAANFR 107
Qy 113 DAL-----GPMLEALDDFFPVYVNLGCVNYFVSENTOYLTRYQDELMKNSVSIHLVGDH 168
Db 108 EGLNLQEGEFLQLQALNGFLVLTALVFASTIDYLGFOQSDVYIHQSVYELHTEDR 167
Qy 169 TEFKNLPLKPSIVNGSGNSGPEPRRNSHFNCRLVPLRDEEGHNDQEAHQRYEQW 228
Db 168 AEFQROL-----HVALNSQ-----CTESGGIGIEATGLPQYV 201
Qy 229 CFVVSQPSIKIEGEDL-QSLICVARRVPMKERPVLPSSSEFTTRDLOGKITSL--- 283
Db 202 CY--NPQIIPRENSPLMERFCIC-----RLRCLDSSGFLA--MNFQKLYLHGOK 250
Qy 284 ---DTS-----TYRAAKRGMEDLVRCIOKFNQHE----- 312
Db 251 KKGKDGSLPQALFAITPLOPPSILEIRTKNEIFFTKHKLDFTPIGCDAKGRIVLGY 310
Qy 313 -----GESVYAKRHHHEVLHQGLAFQSIQYRFSLSDTGLVAQTKSKLI 356
Db 311 TEAEICTRGSGYGFTHAADMLYCAESHIRMTIGSGMIVFRLTKNRMTVWQSNAR-- 369
Qy 357 RSQTTNEPOLVYSLHMLHREQVNCVMPNDLTGOTMGKPLNPISNSPAHQALSGNPGOD 416
Db 370 -----LYKNGRPDYIIYQRLDEBETENLRKNTLP-----FMFTTGEA 411
Qy 417 MTLSSNINFP-INGPKQKGMPMGRFGSGGNNHVSQMATTPOGSNTALKMNSPOSQSP 475
Db 412 VLYEATNPFPAIMDP-----LPLRTKNGTSGKD-----SATYSTLSKSL- 451
Qy 476 GMPGQPTSMLSPRHRMSPGVAGSPRIIPSPQSPAGSLSPVGVCSGNSHSYNTSSLN 535
Db 452 -----NPSLSLAAMQODESITLYP-----ASTYSTAPFENNFN 487
Qy 536 ALQALSEGHSVLSGSSIASPDLKMGNLQNSPVNMNPPLSKMGSLSDKDFGLYGEPSG 595
Db 488 --ESMNE-----CRNQDNTAPMGNDITLKHEQID-----QPD- 519
Qy 596 TTGOAESCHPEQKETNDPNLPPAVSSERADQSRIL-----DSKQTKLL----- 642
Db 520 --VNSFAGGHPQLFQDSKNSDLYSIKMLGIDFEDIRHMQNEKFFRNDPFGSVEFRIDL 577
Qy 643 --QLLTSSDQMEPSPLASSLSDTKKDSGSLPGSGSTHGTSYLKXKHILRLLODSSSP 700
Db 578 TDEILTYVODSLSKSPFIP--SDYQOQOSLAL-----NSSCMVDEHLHLEQOQHOKQ 629
Qy 701 VDLAKLTAEANGKDLSESSSTAPGSEVTIKOEVPYSPKK-----ENALLRYL 748
Db 630 V-----VEEPQQLCQKMKHMQVNGMEFNMNSNOIYFEN 663
Qy 749 LDKDDTKOI-----GLPEITPKLERLDS-----KTDRPSNKLIAKTEKEEM-----S 792
Db 664 CPOQDPQOYVNTFDLHGISOEFYKSEDMSPYTONFTSCNPOVLPJHSHKTELDYPMGS 723
Qy 793 FEPGQPSGLDNLEETLDLQ-----NSQLPOLPEPTRRPAAPAGVQKAIINDL 843
Db 724 FEPSPYPTT--SSLEDFVTCQLPRNQKHGLPQSAITTPQV--CYAGAV-----SM 771
Qy 844 MOLTAE-----NSPVTPVGAQKTALRISQSTFNNPRGQGLRLLPQNLPLDITL 893
Db 772 YQCPPEPQHTHVQGMQYVPLR--GQAFALNFKQNGVLNMEYTPAELENNINNTQTTHQLPL 830
Qy 894 OSPTGAGFPFPIRNS 908
Db 831 HHPSEARFPFDLTSS 845

```

Search completed: September 7, 2002, 10:38:25  
Job time: 287 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 12, 2002, 03:35:41 : Search time 6106.83 Seconds  
(without alignments)  
13605.628 Million cell updates/sec

Title: US-09-842-256-1

Perfect score: 6156  
Sequence: 1 GCGCGCCGACGCTCGGCTA.....CCACTTTATAGTTGTTTTT 6156

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

EST.\*  
1: em\_estbda.\*  
2: em\_esthum.\*  
3: em\_estin.\*  
4: em\_estnu.\*  
5: em\_estov.\*  
6: em\_estpl.\*  
7: em\_estro.\*  
8: em\_hlc.\*  
9: gb\_estl.\*  
10: gb\_est2.\*  
11: gb\_hlc.\*  
12: gb\_gss.\*  
13: em\_gss\_hum.\*  
14: em\_gss\_inv.\*  
15: em\_gss\_pln.\*  
16: em\_gss\_vrt.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	839.2	13.6	1003	10	BM459228
2	746.8	12.1	918	10	BM451917
3	698	11.3	891	10	BM462410
4	674.2	11.0	864	10	BM1734567
5	639.4	10.4	1070	10	BM1248968
6	634	10.3	742	9	BM114474
7	553	9.0	716	10	BM156653
8	548.6	8.9	618	9	BM854255
9	543.8	8.8	941	10	BM1694248
10	534.2	8.7	610	10	BMF044623
11	520.8	8.5	524	9	BM7440705
12	519.6	8.4	555	9	BM170360
13	493.2	8.0	631	9	BM650290
14	492.4	8.0	566	10	BM987973
15	481.2	8.0	545	9	BM031423
16	487.6	7.9	543	9	BM031427
17	483.6	7.9	655	9	BM628347

18	483	7.8	559	10	BM835556
19	480.8	7.8	542	9	BM465013
20	480	7.8	484	9	BM298513
21	478.2	7.8	535	10	BM090425
22	473.6	7.7	777	10	BM172538
23	464.6	7.5	592	10	BM990265
24	462.2	7.5	467	9	BM7444255
25	446.8	7.3	466	9	BM504723
26	445.2	7.2	906	10	BM168417
27	444.2	7.2	556	9	BM666871
28	443	7.2	629	9	BM658506
29	439.4	7.1	673	10	BM923396
30	437.2	7.1	564	9	BM170504
31	434.2	7.1	688	9	BM65333
32	433.4	7.0	435	9	BM127080
33	424	6.9	537	10	BM350947
34	423.8	6.9	440	9	BM165489
35	423.2	6.9	580	9	BM258528
36	422.2	6.9	553	10	BM418851
37	420.8	6.8	448	9	BM027474
38	413	6.7	413	9	BM393449
39	408.8	6.6	485	10	BM651680
40	399.6	6.5	410	9	BM169694
41	397.6	6.5	619	9	BM1596748
42	396.8	6.4	401	9	BM994479
43	395.8	6.4	665	9	BM613358
44	382.4	6.2	384	9	BM185018
45	378	6.1	681	9	BM640117

## ALIGNMENTS

RESULT 1  
LOCUS BM459228 1003 bp mRNA linear EST 05-FEB-2002  
DEFINITION AGENCOURT\_641526 NIH\_MGC\_85 Homo sapiens cDNA clone IMAGE:5495761  
5', mRNA sequence.  
ACCESSION BM459228  
VERSION BM459228.1 GI:18508268  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 1003)  
AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
Tissue Procurement: Lou Staudt  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at:  
<http://image.lnl.gov>  
Plate: L1AM1213 row: d column: 02  
High quality sequence stop: 700.

## FEATURES

1..1003  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5495761"  
/clone\_id="NIH\_MGC\_85"  
/tissue\_type="lymphoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: lymph. Vector: PCMV-SPORE; Site:1: NotI; Site:2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.867 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH\_MGC Library."

BASE COUNT	289 a	298 c	236 g	178 t	2 others
ORIGIN					
Query Match		13.6%	Score 839.2	DB 10	Length 1003
Best Local Similarity		94.1%	Pred. No. 4.5e-187		
Matches 871; Conservative		0	Mismatches 35	Indels 0	Gaps 0

OY	3494	CAGTGGATCCACAAAGCTTCTCAAGTCAGAGATTTCCAAATTCATGCTGGGAGCAAGAAAGGCC	3553
Db	1	CAGTAGATCCAAACAGTTCTCTAAGTCAGAGATTTCCAAATTCATGCTGGGAGCAAGAAAGGCC	60
OY	3554	CCGTTTTCCACAGCAGTAGTAGCATCTCAGGCACAAATGGCCACAGGGTAGCTATTCTCCCA	3613
Db	61	CCGTTTTCCACAGCAGTAGTAGCATCTCAGGCACAAATGGCCACAGGGTAGCTATTCTCCCA	120
OY	3614	TGCAGATCCAAACTTTCACACCATATGGGACACAGCGGCTTAGTTATGCGACACTCCGATATGC	3673
Db	121	TGCAGATCCAAACTTTCACACCATATGGGACACAGCGGCTTAGTTATGCGACACTCCGATATGC	180
OY	3674	AGCCCAAGACCGGGCCTCAGGGCCACGGGCGCTTAGTGAGAACACAGCCAAATTCAACTAAGAC	3733
Db	181	AGCCCAAGACCGGGCCTCAGGGCCACGGGCGCTTAGTGAGAACACAGCCAAATTCAACTAAGAC	240
OY	3734	TTCAACTTCAGCATGCGCTTCCAAGCACAGACAGAAATCGCCAGCCACTTATGAATCAAAATCA	3793
Db	241	TTCAACTTCAGCATGCGCTTCCAAGCACAGACAGAAATCGCCAGCCACTTATGAATCAAAATCA	300
OY	3794	GGAATGTTCCAAATGTGAATCTTGACTCTTAGAGCCTTAGAGTACCACACAGGCACTATTA	3853
Db	301	GGAATGTTCCAAATGTGAATCTTGACTCTTAGAGCCTTAGAGTACCACACAGGCACTATTA	360
OY	3854	ATGCACAGATGCTGGGCCAGAGACAGAGAGGGAATCCTGAACAGCATCTTCACACAGAC	3913
Db	361	ATGCACAGATGCTGGGCCAGAGACAGAGAGGGAATCCTGAACAGCATCTTCACACAGAC	420
OY	3914	AAATGCATCAGACACAGCAAGTTTCAGCAACGAATTTGATGATGAGAGACAGAGGGTTGA	3973
Db	421	AAATGCATCAGACACAGCAAGTTTCAGCAACGAATTTGATGATGAGAGACAGAGGGTTGA	480
OY	3974	AATATGACACCAAGCATGCTGGCTCCTAGTGGTATATGTCACACATATGACCAACCTCCGA	4033
Db	481	AATATGACACCAAGCATGCTGGCTCCTAGTGGTATATGTCACACATATGACCAACCTCCGA	540
OY	4034	TTTCCCAAGCAATGCACAGCAAGTTTCATTTCTCTCAACACTAAGGAATGAATGTCAGCAAC	4093
Db	541	TTTCCCAAGCAATGCACAGCAAGTTTCATTTCTCTCAACACTAAGGAATGAATGTCAGCAAC	600
OY	4094	CTGATTCACAGGCTTTACTGGGGCTACGACTCCCCAGAGCCCACTTATGTCACCCCGAATGG	4153
Db	601	CTGATTCACAGGCTTTACTGGGGCTACGACTCCCCAGAGCCCACTTATGTCACCCCGAATGG	660
OY	4154	CACATATCACAGAGTCCCATGATGTGAACAATGTCAGGCCCAACCCAGCCTATTCAGGCCCCCT	4213
Db	661	CACATATCACAGAGTCCCATGATGTGAACAATGTCAGGCCCAACCCAGCCTATTCAGGCCCCCT	720
OY	4214	CCGACATTAATAGATGGGGGAGGAGGGAATGAGGCGGAACAGCATGTTTCCACAGCAT	4273
Db	721	CCGACATTAATAGATGGGGGAGGAGGGAATGAGGCGGAACAGCATGTTTCCACAGCAT	780
OY	4274	CCCCACACACATTTTGGGCGACAGCAACACACAGCATGTATGACAGTAACAATGAACATCA	4333
Db	781	CCCCACACACATTTTGGGCGAGCAAGCAACACACAGCATGTATGACAGTAACAATGAACATCA	840
OY	4334	ATGTGTCCATGGCGACCAACACAGGTGGCATGACAGCATGGAACCAAGATGACAGAGACAA	4393
Db	841	ATGTGTCCATGGCGACCAACACAGGTGGCATGAGCCAGATGGAACCAAGATGACAGAGACAA	900
OY	4394	TCAGCATGACCTCAGTGAACCTCCGTTG 4419	
Db	901	GGACAGAGATCAGCATGTGACCTCAGTG 926	

[illegible]



Df	361	TCAATGCATCGAAGCCCTCAGCGAGGGGCGCATGGGGTCTCACTCGGGGTCTTCGCTGGGCTT	420
OY	1823	CACGAGACTTAATAATGGGCATTTTGCAAACCTCCCCAGTAAATATGAATCTCCCCAC	1882
Df	421	CACCGGACCTAAAAATGGGCAATTGGCAAACTCCCAGTTAAATATGAATCTCCCCAC	480
OY	1883	TCAGCAATATGGGAAGCTTGGACATCAAAAAGACTGTTTTGACTATATGAGGAGGCCCTTG	1942
Df	481	TCAGCAATATGGGAAGCTTGGACATCAAAAAGACTGTTTTGACTATATGAGGAGGCCCTTG	540
OY	1943	AAGTACAATCTGGACAGACAGACAGCTGCGATCTCTGAGAGCAAAAGAAACAATG	2002
Df	541	AAGTACAATCTGGACAGACAGAGGCCACTGCGATCTCTGAGAGCAAAAGAGGCCAATG	600
OY	2003	ACCCTAACCTGCCCCCGGCGCTGAGCACTGAGAGAGCTGAGGGGAGAGAGACTGCTGATG	2062
Df	601	ATTCCGATATGCCCGAGGCGCGCACGGGGGAGAGGGGCTGAGGAGACACAGCCGGCTGATG	660
OY	2063	ACAGCAAAAGGGGAGACCAAACTCTGACAGCTGCTACACCACCAATCTGA-TCAGATGAG	2121
Df	661	ACAGCAAAAGGGGAGACCAAACTCTGACAGCTGCTACACCACCAATCTGA-TCAGATGAG	720
OY	2122	CCCTTGGCCCTTAG-CCAGCTCTTTTGTGCGATAC-AAACAAGACTCCACAGGTAGCTTGC	2179
Df	721	CCCTTACACCTTTTGGCCAGCTCTCTGTGCGAGACAAACAGAGACTCAACAGGAGACTTGC	780
OY	2180	CTGGTTC-TGGGCTTACACATGGAACCTGCTCAAGAAGAACATTAATTTTGCACAGA	2238
Df	781	CTGGGCTTGGGCTTGGCCAGCTGCTGCTCAAGAAGAACATTAATTTTGCACAGA	840
OY	2239	CTCTTGCAGGACAGACAGTTCCCC	2261
Df	841	CTTCTTACGACAGACAGTTCOC	863
RESULT	5		
LOCUS	B1248968	1070 bp	mRNA linear EST 17-JUN-2001
DEFINITION	60299257221 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:5148545	5'	
ACCESSION	B1248968		
VERSION	B1248968		
KEYWORDS	EST.		
SOURCE	house mouse.		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.		
TITLE	1 (bases 1 to 1070)		
JOURNAL	NIH-MGC http://mgc.nci.nih.gov/		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: cga@bbs.fda.hhs.gov		
	Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys		
	cDNA Library Preparation: Life Technologies, Inc.		
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)		
	DNA Sequencing by: Incyte Genomics, Inc.		
	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at:		
	http://image.lnl.gov		
	Plate: L1AM1366 row: p column: 18		
	High quality sequence stop: 905.		
FEATURES	Location/Qualifiers		
source	1..1070		
	/organism="Mus musculus"		
	/strain="C57/B6"		
	/db_xref="taxon:10090"		
	/clone_image="5148545"		
	/clone_lib="NCI CGAP Mam5"		
	/tissue_type="tumor, gross tissue"		
	/dev_stage="7 months"		
	/lab_host="DH10B"		
	/note="Organ: mammary; Vector: pCMV-Sport6; Site_1: SalI;		





QY 4945 TCTTGGCTTTGTTTCCCTGGCTAACAGCTAGTCCCAAGATTATGATTTATCTGGGG 5004  
|||||  
Db 361 TCTTGGCTTTGTTTCCCTGGCTAACAGCTAGTCCCAAGATTATGATTTATCTGGGG 420  
QY 5005 AAGAAAAAGATTTTAAAAAATTAACTAAGATGTTTAAAGCTTAAGCCCTGAATTTG 5064  
|||||  
Db 421 AAGAAAAAGATTTTAAAAAATTAACTAAGATGTTTAAAGCTTAAGCCCTGAATTTG 480  
QY 5065 GGATGGACAGACAGACACCGCTGATATTATACAGACACCCAGTCCGTG 5124  
|||||  
Db 481 GGATGGACAGACAGACACCGCTGATATTATACAGACACCCAGTCCGTG 540  
QY 5125 AAGACCAACAAAGTCAACAGTCTATCTAGAAAGCTTAAGACCATGTTGGAAGAGT 5184  
|||||  
Db 541 AAGACCAACAAAGTCAACAGTCTATCTAGAAAGCTTAAGACCATGTTGGAAGAGT 600  
QY 5185 CTCACATTTACTGAAAGATGAAAGAGCCTGAGAGGGCTTTAATACATTAATAT 5244  
|||||  
Db 601 CTCACATTTACTGAAAGATGAAAGAGCCTGAGAGGGCTTTAATACATTAATAT 660  
QY 5245 TTTTTCCTGTTTCTTCTT 5263  
|||||  
Db 661 TTTTTCCTGTTTCTTCTT 679

RESULT 7  
B1156653 716 bp mRNA linear EST 05-JUL-2001  
LOCUS 60292125F1 NIH\_CGAP\_Mam3 Mus musculus cDNA clone IMAGE:5061685 5',  
DEFINITION mRNA sequence.  
ACCESSION B1156653  
VERSION B1156653.1 GI:14616654  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 716)  
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: Lohar Hennighausen Ph.D., Chu-Xia Deng Ph.D.  
CDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
plate: L1AM11166 row: m column: 14  
High quality sequence stop: 714.

FEATURES  
source  
1..716  
Location/Qualifiers  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="IMAGE:5061685"  
/clone\_lib="NIH\_CGAP\_Mam3"  
/tissue\_type="tumor, gross tissue"  
/lab\_host="DH10B"  
/note="Organ: mammary; Vector: pCMV-SPORT6; Site:1: NotI;  
Site:2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert 2 kb. Library constructed by Life  
Technologies, catalog #12017-018. Investigators providing  
samples: Lohar Hennighausen/Chu-Xia Deng, NIH Reference  
for transgenic model: Xu et al., Nature Genetics 22, 37-43  
(1999). Note: this is a NCI/CGAP Library."

BASE COUNT 189 a 222 c 177 g 128 t  
ORIGIN

Query Match 9.0%; Score 553; DB 10; Length 716;  
Best Local Similarity 87.3%; Pred. No. 1.7e-119;

Matches 631; Conservative 0; Mismatches 85; Indels 7; Gaps 2;  
QY 3727 CTAGACCTTCAACTTGGATGAGTCGCTCAAGCAGACAGAAATGCCAGCATATGAAT 3786  
|||||  
Db 1 CTAGACCTTCAACTTGGATGAGTCGCTCAAGCAGACAGAAATGCCAGCATATGAAT 60  
QY 3787 CAAATCAGCAATGTTTCCAAATGGAATTGACTGTGAGGCTGTGATACCAACAGGCA 3846  
|||||  
Db 61 CAAATCAGCAATGTTTCCAAATGGAATTGACTGTGAGGCTGTGATACCAACAGGCT 120  
QY 3847 CCTATTAAATGACAGATGCTGGCCAGAGACAGAGGAAATCTGACAGATCTGCA 3906  
|||||  
Db 121 CCTATTAAATGACAGATGCTGGCCAGAGACAGAGGAAATCTGACAGATCTGCA 180  
QY 3907 CAGAGCAAAATGATCAGCAACAGCAAGTTTCAGCAAGCACTTTGATGAGAGGACAA 3966  
|||||  
Db 181 CAGAGCAAGT-----GCAGCAGCAGGTGACAGCAGGACTGTGATGATGAGAGCAG 234  
QY 3967 GGGTGAATATGACACCAAGCATGTGGCTCTAGTGTATGACCAACTATAGCAAC 4026  
|||||  
Db 235 GGGTGAATATGACACCAAGCATGTGGCTCTAGTGTATGACCAACTATAGCAAC 294  
QY 4027 CCTGGATTTCCCGAGCAAAATGACAGAGTTTCCATTTCCCTCAACATAGGCAATAGT 4086  
|||||  
Db 295 CCTGGATTTCCCGAGCAAAATGACAGAGTTTCCATTTCCCTCAACATAGGCAATAGT 354  
QY 4087 CAGCAACCTGATCAGCAGCTTACTGTGGGTACAGACTCCCGAGAGCCATATATGACCC 4146  
|||||  
Db 355 CAGCAACCTGATCAGCTTACTGTGGGTACAGACTCCCGAGAGCCATATATGACCC 414  
QY 4147 CGAATGGACATATACAGAGAGTCCCATGATGCAACAGTCTCAGGCCAACCCAGCTATCAC 4206  
|||||  
Db 415 CGAATGGACATATCAGAGAGTCCCATGATGCAACAGTCTCAGGCCAACCCAGCTATCAC 474  
QY 4207 GCGGCTCCGACATTAATGATGGATGGGGGAGGAGCAATGGGGGCAACAGCATTTTCC 4266  
|||||  
Db 475 GCGGCTCCGACATTAATGATGGATGGGGGAGGAGCAATGGGGGCAACAGCATTTTCC 534  
QY 4267 CAGCAGTCCCGACACACATTTGGGCGCAAGCAACAGCAGATGTATACCAACATG 4326  
|||||  
Db 535 CAGCAGTCCCGACACACATTTGGGCGCAACAGCAGATGTATACCAACATG 594  
QY 4327 AACATCAATGTGTCATGGCGACCAACAGAGTGGCATGACAGCATGAACAGATGACA 4386  
|||||  
Db 595 AACATCAATGTGTCATGGCGACCAACAGAGTGGCATGACAGCATGAACAGATGACA 654  
QY 4387 GGACATATGACATGACATGACATGACATGACATGACATGACATGACATGACATG 4446  
|||||  
Db 655 GGACATATGACATGACATGACATGACATGACATGACATGACATGACATGACATG 713  
QY 4447 CCC 4449  
|||  
Db 714 CCC 716

RESULT 8  
AW854255 618 bp mRNA linear EST 19-MAY-2000  
LOCUS RC3-CT0254-100500-211-d11 CT0254 Homo sapiens cDNA, mRNA sequence.  
DEFINITION  
ACCESSION AW854255  
VERSION AW854255.1 GI:7949948  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens.

REFERENCE 1 (bases 1 to 618)  
AUTHORS Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
Dias Neto, E., Garcia Correa, R., Verjowski-Almeida, S., Briones, M. R.,  
Nagai, M. A., da Silva, M. Jr., Zago, M. A., Bordin, S., Costa, F. F.,  
Goldman, G. H., Carvalho, A. F., Matsukuma, A., Baia, G. S., Simpson, D. H.,  
Brunstein, A., de Oliveira, P. S., Bucher, P., Jongeneel, C. V., O'Hare,  
M. J., Soares, F., Brentani, R. R., Reis, L. F., de Souza, S. J. and  
Simpson, A. J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
MEDLINE 20202663  
COMMENT Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Genome Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2-RC3-CT0254-100500-211-d11c3-2000-05-10&tl=1)  
Seq primer: puc 18 forward  
High quality sequence start: 70  
Location/Qualifiers  
1. 618

FEATURES  
source  
1. 618  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_id="CT0254"  
/dev\_stage="Adult"  
/note="Organ: colon; Vector: puc18; Site:1: Sma1; Site:2: Sma1; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 199 a 116 c 121 g 182 t  
ORIGIN

Query Match 8.9%; Score 548.6; DB 9; Length 618;  
Best Local Similarity 96.4%; Pred. No. 1.8e-118;  
Matches 593; Conservative 0; Mismatches 19; Indels 3; Gaps 3;

QY 4750 ACAGCTGGAAGCTCGCATCCAGACAGTCGCTGCTGCTCCGTCATTCACCTTAGTGC 4809  
DB 5 ACAGCGTGAAGCTCGCATCCAGACAGTCGCTGCTGCTCCGTCATTCACCTTAGTGC 63  
QY 4810 AACTTAAATCTCTCTCCGCAAGTAATGTGACAGGCCATTCATACCATGTCAGAT 4869  
DB 64 AACTTAAATCTCTCTCCGCAAGTAATGTGACAGGCCATTCATACCATGTCAGAT 123  
QY 4870 TGAATGATTTAAATGATATGATTTAAAG-GAAGACATGCTCTGTTCTGTTCTGTTGCG 4928  
DB 124 TGAATGATTTAAATGATATGATTTAAAG-GAAGACATGCTCTGTTCTGTTCTGTTGCG 183  
QY 4929 GTTCCAGACACTGCTTCTGCTTGTGTTTCCGCTAGCAGTCTAGTGCCTAAGATTA 4988  
DB 184 GTTCCAGACACTGCTTCTGCTTGTGTTTCCGCTAGCAGTCTAGTGCCTAAGATTA 243  
QY 4989 AGATTTATCTGGG-GGAAGAGAAAAGATTTTAAATAAATTAACCTAAGATGTTTAA 5047  
DB 244 AGATTTATCTGGGAGAGAGAAAAGATTTTAAATAAATTAACCTAAGATGTTTAA 303  
QY 5048 GCTAAACCTGAATTTGGATGGAAGACAGACACCCGTGACAGCGCTGATTTTACA 5107  
DB 304 GCTAAACCTGAATTTGGATGGAAGACAGACACCCGTGACAGCGCTGATTTTACA 363  
QY 5108 GACACACCCAGTGGGTGAAGACCAAAAGTCACAGTGTCTCTAGAAAAGCTCTAAG 5167  
DB 364 GACACACCCAGTGGGTGAAGACCAAAAGTCACAGTGTCTCTAGAAAAGCTCTAAG 423  
QY 5168 ACCATGTTGAGAAAGTCTCCAGTTTACGACAGATGAAAGAGAGCCGTAGAGGGGTG 5227  
DB 424 ACCATGTTGAGAAAGTCTCCAGTTTACGACAGATGAAAGAGAGCCGTAGAGGGGTG 483  
QY 5228 TTAACATTAAGCAATATTTTCTGTTTCTTTGTTTAAACCAAACTGTTCACT 5287

DB 484 TTAACATTAAGCAATATTTTCTGTTTCTTTGTTTAAACCAAACTGTTCACT 543  
QY 5288 GAATCATGAATTTGAGAGAAATTAATTTTCAATTTCTAAATTAAGTCCCTTTAGTTGATC 5347  
DB 544 GAATCATGAATTTGAGAGAAATTAATTTTCAATTTCTAAATTAAGTCCCTTTAGTTGATC 603  
QY 5348 AGACAGCTGAATCA 5362  
DB 604 AGACAGCTGAATCA 618

RESULT 9  
B1694248 941 bp mRNA linear EST 18-SEP-2001  
LOCUS 603347586F1 NCL\_CGAP\_Mam2 Mus musculus cDNA clone IMAGE:5375468 5,  
DEFINITION mRNA sequence.  
ACCESSION B1694248  
VERSION B1694248.1 GI:15656877  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 941)  
NHL-MGC http://mgc.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
AUTHORS Contact: Robert Strausberg, Ph.D.  
Email: c9abs-remail.nih.gov  
Tissue Procurement: Gilbert Smith, Ph.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LMNL at:  
http://image.llnl.gov  
Plate: LMNL1955 row: 0 column: 21  
High quality sequence stop: 800.

FEATURES  
source  
1. 941  
Location/Qualifiers  
/organism="Mus musculus"  
/strain="FVB/N-3"  
/db\_xref="taxon:10090"  
/clone\_image="5375468"  
/clone\_id="NCL\_CGAP\_Mam2"  
/tissue\_type="tumor, biopsy sample"  
/dev\_stage="5 months"  
/lab\_host="DH10B"  
/note="Organ: mammary; Vector: pCMV-SPORT6; Site:1: SalI; Site:2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigator providing samples: Gilbert Smith, NIH"

BASE COUNT 244 a 265 c 251 g 181 t  
ORIGIN

Query Match 8.8%; Score 543.8; DB 10; Length 941;  
Best Local Similarity 80.0%; Pred. No. 2.6e-117;  
Matches 730; Conservative 0; Mismatches 167; Indels 16; Gaps 7;

QY 3780 TATGAATCAATACGACATGTTCCATGTGAACCTGACTGAGGCTGAGTACCAAC 3839  
DB 1 TATGAATCAATACGACATGTTCCATGTGAACCTGACTGAGGCTGAGTACCAAC 58  
QY 3840 ACAGGACCTTATATGACAGATGCTGGCCAGAGACAGAGGGAATCTGACACAGCA 3899  
DB 59 TCAGGCTCCTATTAATGACAGATGCTGGCCAGAGACAGAGGGAATCTGACACAGCA 118  
QY 3900 TCTTTCAGACAGCAATGATCAAGCAAGCAAGTTGACAGCAAGCAATTGATGATGAG 3959  
DB 119 CCTTCGGCAGAGCGACAGAT-----GCAGACAGCAGTGCAGCAGCGACTGATGATGAG 172

QY	3960	AGGACAAAGGGTTGATPATGATGACACCAAGATGAGTGGCTCTCTTATGGTATGACGACAATAT	4019
Db	173	AGGACAAAGGGCTTGAATGTGACCCCAAGCATGTGTGGCTCCCTGGCTTACACGACGCCT	232
QY	4020	GAGCAACACCTCGGATTTCCCCAGGCAAAATGACAGCAGTTTCCATTTCCTCCAACTAGAG	4079
Db	233	GAGCAATCCCCGGATTCCTCCAGGCAATGCCAGATGTCCTCCATTTCCTCCGAACTAGAG	292
QY	4080	AATTAAGTCAGCAGCTGATGACAGGCTTTACTGTGGGCTTACGACTCCCCAGAGCCACTTAT	4139
Db	293	AATAAGTCAGCAACCTGATTCCTGGCTTTACTGTGGGCTTACGACTCCCCAGAGCTCTTAT	352
QY	4140	GTCCACCCCGAATGGGACATACAGACAGTCCCATGATGTGAACAGTCTCCAGGCCAACCCAGC	4199
Db	353	GTCTCCCGGATGGACATATCTCAAGATGCCATGATGACAGCTCTCAAGCCCAACCCAGC	412
QY	4200	CTATCAGGCCCCCTCCGACATTAATGATGGGCGCAGGGGAACATGAGGCGGAACAGAT	4259
Db	413	CTACCAAGCCCACTTACGACATTAATGATGGGCAACAGGGAGCATGGGTGGAAACAGCAT	472
QY	4260	GTTTTCCAGCAGTCCCAACCACTTTGGGCGAGAGCAAGCAACACACAGCATGTACAGTAA	4319
Db	473	GTTTCAACAGCAGTCCCAACCACTTTGGGCGAACAAGCAACACACAGCATGTATAGTAA	532
QY	4320	CACACATGAACATCAATGTGTCCATGTGCGGACACACACAGAGTGGCATGAGCAGCATACCA	4379
Db	533	CACACATGAACATCAATGTGTGATGTGCAACCAACCAAGGTTGGCTTGAGCAGCATACCA	592
QY	4380	GATGACAGACAGATCAGCATGACCTCAGTGCACGTCGCTGCTAGCTGACAGGCTGCTGC	4439
Db	593	GATGACAGGCGCAGATGAGCATGACCTCAGTGCACGTCGCTGCTAGCTGACAGGCTGCTGC	651
QY	4440	CATGGGTCCCGAGCAGGTTAATGATCTGTCTGAGGGGAGGCAAC--TGTTCCCAA	4496
Db	652	CATGGGTCCCGAGCAGGTTAATGATCTGTCTGAGGGGAGGCAACCTTTTCCCAA	711
QY	4497	CCAGCTG--CCTGGAAATGATATGATTAAGCAGGAGGAGACAAAC--ACGAAATATTG	4553
Db	712	CCAACTGGCGCTGGAATGATGATGATCAAGCAAGGAGGAGATCATCTTCCGGAATACG	771
QY	4554	CTGACATGCTGAAACCAAGTTCCTTCTGAGTACCGGGGCTCAGTGGTCAAAACACTT	4613
Db	772	GTTGACCCGGGAAACATGTCTGCAATCTTCTTCAACCACTAGGTTTACAAAACATTTTA	831
QY	4614	CCAGCTGTGAGA--GCTGTGTCTAATTTGTTCAACCACTGACCTGCGAGGTTCTGC	4672
Db	832	CCAGCTGTGAGAGGCTGCGGTTCTTGTGTGAGCAACCTGACATTTGCCGCGAGGTTTCC	891
QY	4673	TGAGCAGACAGAG 4685	
Db	892	AGGAGCATTTACAG 904	
RESULT 10			
EF044623/c			
LOCUS	EF044623	610 bp	mRNA
DEFINITION	BP250020A20G6 Soares normalized bovine Placenta Bos taurus cDNA		EST 10-OCT-2000
ACCESSION	BP044623		
VERSION	BP044623.1	GI:10761678	
KEYWORDS	EST.		
SOURCE	cow.		
ORGANISM	Bos taurus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;		
	Bovidae; Bovinae; Bos.		
	I (bases 1 to 610)		
	Lewin,H.A., Soares,M.B., Rebeiz,M., Pardinas,J., Liu,L. and Larson		
	J.H.		
TITLE	Bovine ESTs		
JOURNAL	Unpublished (2000)		
COMMENT	Contact: Lewin, H. A.		
	W. M. Keck Center for Comparative and Functional Genomics		

|||||  
Db 130 GGGTCTCTTGTGTCGTAACCTGGAAAGCAAGCTGTGTTGTTTCAGACAATG -NNNA 72  
Oy 595 CAGATATCAAGCTATTAACCAAGAGAGCTGATGAACAAAATGATATAGCATCTTGAT 654  
Db 71 CAGTACCTTAAGGTTAAACGAGAGAGCTGATGAACAAAAGTGTATTAAGTCTCTGCAC 12  
Oy 655 GTTGGGACCA 665  
Db 11 GTTGGGACCA 1

RESULT 11  
AA740705 524 bp mRNA linear EST 07-FEB-1998  
LOCUS ny98404.s1 NCI\_CGAP\_GCB1 Homo sapiens CDNA clone IMAGE:1286286 3,  
DEFINITION similar to TR:Q15356 Q15396 TRANSCRIPTONAL INTERMEDIARY FACTOR 2,  
; mRNA sequence.  
ACCESSION AA740705  
VERSION AA740705.1 GI:2779297  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 524)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP).  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgaps-r@mail.nih.gov  
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,  
Ph.D., Gerald Marti, M.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
Bonaldo, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LML at:  
[www.bio.lml.gov/bdrrp/image/image.html](http://www.bio.lml.gov/bdrrp/image/image.html)  
Insert Length: 661 Std Error: 0.00  
Seq primer: -40m13 fwd. RT from Amersham  
High quality sequence stop: 367.  
location/Qualifiers  
1.524  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1286286"  
/clone\_1id="NCI\_CGAP\_GCB1"  
/tissue\_type="germinal center B cell"  
/lab\_host="DH10B"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker. Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
was prepared from human tonsillar cells enriched for  
germinal center B cells by flow sorting (CD20+, IgD-),  
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman  
(NCI) and Dr. Gerald Marti (CEBR). CDNA synthesis was  
primed with a Not I - oligo(dT) primer  
15'-TGTTACCAATCTGAAGTGGAGGCGGCTCAATTTTTTTTTTTT-3'  
1. Double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pT73 vector. Library  
went through one round of normalization, and was  
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 171 a 119 c 130 g 104 t  
ORIGIN

Query Match 8.5% Score 520.8; DB 9; Length 524;  
Best Local Similarity 99.6% Pred. No. 6.2e-112;  
Matches 522; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 2147 CGATATCAACAAAGACTCCACAGTAGCTTCCCTGGTTGCGGCTACACATGAACCT 2206  
Db 1 CGGATACAAACAAAGACTCCACAGTAGCTTCCCTGGTTGCGGCTACACATGAACCT 60  
Oy 2207 CGCTCAAGAGAGACATAAATTTTGCACAGACTCTTGCAGAGACAGCTTCCCTGTGG 2266  
Db 61 CGCTCAAGAGAGACATAAATTTTGCACAGACTCTTGCAGAGACAGCTTCCCTGTGG 120  
Oy 2267 ACTTGGCCCAATTAACAGACACAGCCACAGCAAGACCTAAGCAGAGTCCAGCAGCA 2326  
Db 121 ACTTGGCCCAATTAACAGACACAGCCACAGCAAGACCTAAGCAGAGTCCAGCAGCA 180  
Oy 2327 CAGCTCCTGATCAGAACTGACTATTAACAGAGCCGCTGAGCCCAAGAGAGAGGA 2386  
Db 181 CAGCTCCTGATCAGAACTGACTATTAACAGAGCCGCTGAGCCCAAGAGAGAGGA 240  
Oy 2387 ATGCACCTACTTGGCTATTTGCTAGATAAAGATGATTAAGATATGTTTACAGAAA 2446  
Db 241 ATGCACCTACTTGGCTATTTGCTAGATAAAGATGATTAAGATATGTTTACAGAAA 300  
Oy 2447 TAACCCCAAACTGAGAGACTGAGACAGTAAAGACATCTCCAGTAACCAAAATTA 2506  
Db 301 TAACCCCAAACTGAGAGACTGAGACAGTAAAGACATCTCCAGTAACCAAAATTA 360  
Oy 2507 TAGCAATGAAGACTGAGAGAGAGAGAGATGAGCTTGGAGCTGAGCAGCTGCAAGT 2566  
Db 361 TAGCAATGAAGACTGAGAGAGAGAGAGATGAGCTTGGAGCTGAGCAGCTGCAAGT 420  
Oy 2567 AGCTGGACAACTGGAGAGAGATTTGGATGATTTGCAGAAATGATTAACAGAGCTTT 2626  
Db 421 AGCTGGACAACTGGAGAGAGATTTGGATGATTTGCAGAAATGATTAACAGAGCTTT 480  
Oy 2627 TCCCAAGACACAGAGCCAGGCGCCCTGCTGATCAGTTGACAG 2670  
Db 481 TCCCAAGACACAGAGCCAGGCGCCCTGCTGATCAGTTGACAG 524

RESULT 12  
BE170360  
LOCUS BE170360 555 bp mRNA linear EST 21-JUN-2000  
DEFINITION OV4-HT0538-020300-123-e04 HT0538 Homo sapiens CDNA, mRNA sequence.  
ACCESSION BE170360  
VERSION BE170360.1 GI:8633081  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 555)  
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,  
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,  
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
Simpson,A.J.  
Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
CONTACT: Simpson A.J.G.  
20202663  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: [asimpson@ludwig.org.br](mailto:asimpson@ludwig.org.br)  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(<http://www.ludwig.org.br/scripts/gethtml2.pl?fl=674-QV4-HT0538-020>)  
300-123-e046t3-2000-03-026t4-1)  
Seq primer: puc 18 forward  
High quality sequence stop: 555.



Best Local Similarity 88.0%; Pred. No. 2.1e-105;  
Matches 549; Conservative 0; Mismatches 73; Indels 2; Gaps 1;

```

Qy 1 GCGGCGCGACCTCGGCTACAGTTCGGCGCGCAAGTCAAGCCCGACGCGACCGCA 60
Db 10 GCGGCGCGACCTCGGCTACAGTTCGGCGCGCAAGTCAAGCCCGACGCGACCGCA 69
Qy 61 CCTGAGCGCGTGAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120
Db 70 CCTGAGCGCGTGAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 127
Qy 121 TTCTGACCTGTTTACAGCAGAGTTCGTATGTATGTTCAGATGAGTGGGATGGAGAA 180
Db 128 TTCTGACCTGTTTACAGCAGAGTTCGTATGTATGTTCAGATGAGTGGGATGGAGAA 187
Qy 181 AATACCTCTGACCTCTCCAGGCGAGAGCAAGAAAGCGCAAGATCTCTGACCAACTT 240
Db 188 AACACCTCTGACCTCTCCAGGCGAGAGCAAGAAAGCGCAAGATCTCTCCGACGACTC 247
Qy 241 GGAACCCAGCCCGCAAGAAAGCAAGTGAATGCTGAAGCAAGAAATTAATATATA 300
Db 248 GGAACCCAGCCCGCAAGAAAGCAAGTGAATGCTGAAGCAAGAAATTAATATATA 307
Qy 301 GAAGAACTTCGAGAGTGTATTTTGAATTTTGAATATAGCAACTTAACTTCAAA 360
Db 308 GAGAGAGCTGGCGGAGTGTATCTTCGCAAACTTAAATGATATGCAACTTCAACTTCAA 367
Qy 361 CCTGACAAATGTGCAATCTTAAAGAAAGCTGTGACCAATCTCTGACATCAAGCAAA 420
Db 368 CCTGACAAATGTGCAATCTTAAAGAAAGCTGTGACCAATCTCTGACATCAAGCAAA 427
Qy 421 GAGAAACGACAGCTGCAACATGATGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 480
Db 428 GAGAAACGACAGCTGCAACATGATGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 487
Qy 481 CAGGCTGTCATCGACAAGATGCGCGCTGATGCTTGAAGCCCTTGATGGGTTTC 540
Db 488 CAGGCTGTCATCGACAAGATGCGCGCTGATGCTTGAAGCCCTTGATGGGTTTC 547
Qy 541 TTCTTTTATGTAACCTCGAAGGCAAGCTGTGTGTGTGTGACAGAAATGTGACAGATAT 600
Db 548 TTCTTTTATGTAACCTCGAAGGCAAGCTGTGTGTGTGTGACAGAAATGTGACAGATAT 607
Qy 601 CTAAGGTATTAACCAAGAGAGCTG 624
Db 608 CTACGGTATTAACCAAGAGAGCTG 631

```

## RESULT 14

LOCUS BG987973 566 bp mRNA linear EST 13-JUN-2001  
DEFINITION MR2-HT1162-110101-001-b05 HT1162 Homo sapiens cDNA, mRNA sequence.  
ACCESSION BG987973  
VERSION BG987973.1 GI:14392043  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 566)  
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,  
Nagel,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,P.F.,  
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,  
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,  
M.U., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
Simpson,A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
MEDLINE 20202663  
COMMENT Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-27049922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=MR2&cl=MR2-HT1162-  
110101-001-b05&cl=2001-01-11&cl=1)  
Seq primer: puc 18 forward  
High quality sequence stop: 281.  
Location/Qualifiers

## FEATURES

## source

1..566  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="HT1162"  
/dev\_stage="adult"  
/note="Organ: head,neck; Vector: puc18; Site.1: Smar;  
Site.2: Smar; A mini-library was made by cloning products  
derived from ORESMES PCR (U.S. Letters Patent application  
No.196,716 - Ludwig Institute for Cancer Research)  
profiles into the puc 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."

BASE COUNT 166 a 147 c 174 g 79 t  
ORIGIN

## Query Match

Best Local Similarity 94.2%; Score 492.4; DB 10; Length 566;  
Matches 533; Conservative 0; Mismatches 31; Indels 2; Gaps 2;

```

Qy 1805 TAGGGTATGTTGGCTTACACAGCTTAAATGGCAATTTGCAAACTCCCACTTA 1864
Db 1 TAGGGTATGTTGGCTTACACAGCTTAAATGGCAATTTGCAAACTCCCACTTA 60
Qy 1865 ATATGAATTCCTCCCTCAGCAAGATGGAGCTTGAAGCTGCAAAAGCTGTTTGAC 1924
Db 61 ATATGAATTCCTCCCTCAGCAAGATGGAGCTTGAAGCTGCAAAAGCTGTTTGAC 120
Qy 1925 TATATGGGAGCCCTCTGAAGTACAACTGACAAAGAGAGAGAGAGAGAGAGAGAG 1984
Db 121 TATATGGGAGCCCTCTGAAGTACAACTGACAAAGAGAGAGAGAGAGAGAGAGAG 180
Qy 1985 AGCAAAAGAAACAA-TGACCCCAACCTGCCCCCGCGCTGAGCACTGAGAGCTGAC 2043
Db 181 AGCAAAAGAAACAA-TGACCCCAACCTGCCCCCGCGCTGAGCACTGAGAGCTGAC 240
Qy 2044 GGGCAGAGCAGAGCTGCA-TGACAGCAAAAGGAGAGCAAACTCCTGAGCTGCGACAC 2102
Db 241 GGGCAGAGCAGAGCTGCA-TGACAGCAAAAGGAGAGCAAACTCCTGAGCTGCGACAC 300
Qy 2103 CAATCTGATCAGATGAGAGCCCTGCGCTTGAAGCAGCTTTGTGCGATCAAAACAAGA 2162
Db 301 CAATCTGATCAGATGAGAGCCCTGCGCTTGAAGCAGCTTTGTGCGATCAAAACAAGA 360
Qy 2163 CTCACAGTGAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2222
Db 361 CTCACAGTGAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
Qy 2223 TAAATTTTGCACAGATCTGTCGAGCAGCAGTCCCTGCTGAGTGGCCAAAGTTAAC 2282
Db 421 TAAATTTTGCACAGATCTGTCGAGCAGCAGTCCCTGCTGAGTGGCCAAAGTTAAC 480
Qy 2283 AGCAGAAAGCAGAGCAGCAGTGAAGCTGAGAGTCCAGAGCAGCAGCAGTCTGATGAGA 2342
Db 481 AGCAGAAAGCAGAGCAGCAGTGAAGCTGAGAGTCCAGAGCAGCAGCAGTCTGATGAGA 540
Qy 2343 AGTGACTATTAAACAGAGCCGGTGA 2368
Db 541 AGTGACTATTAAACAGAGCCGGGGA 566

```

```

RESULT 15
BE031423          545 bp  mRNA  linear  EST 09-JUL-2000
LOCUS             BE031423
DEFINITION        129986 MARC 1PIG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION         BE031423
VERSION           BE031423.1 GI:8326432
KEYWORDS          EST.
SOURCE            pig.
ORGANISM          Sus scrofa
                  Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
                  Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE         1 (bases 1 to 545)
                  Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
                  Stone,R.T., Heaton,M.P., Grose,W.M., Bennett,G.A., Laegreid,W.W.,
                  and Keele,J.W.
                  Design and use of two pooled tissue normalized cDNA libraries for
                  EST discovery in swine
JOURNAL           Unpublished (2000)
COMMENT           Contact: Smith TPL
                  USDA, ARS, US Meat Animal Research Center
                  PO Box 166, Clay Center, NE 68933-0166, USA
                  Tel: 402 762 4366
                  Fax: 402 762 4390
                  Email: smithemail.marc.usda.gov
                  Single pass sequencing. Bases called and alt-trimmed with phred
                  v0.980904.e. Vector identified by cross_match with the -minscore 18
                  and -mismatch 12 options.
PCR Primers      FORWARD: AGGAACAGCTATGACCAT
                  BACKWARD: GTTTCCCACTCAGCAGC
                  Plate: 66 row: K column: 9
                  Seq primer: ATTTAGTGACACTATAG.
                  Location/Qualifiers
FEATURES          source
                  1..545
                  /organism="Sus scrofa"
                  /db_xref="taxon:9823"
                  /clone_lib="MARC 1PIG"
                  /tissue_type="pooled"
                  /lab_host="DH10B"
                  /note="Vector: pCMV SPORT6; Site.1: XbaI; Site.2: XhoI;
                  library made from pooled tissue from day 11, 13, 15, 20,
                  and 30 embryos."
BASE COUNT       170 a 110 c 140 g 125 t
ORIGIN
Query Match      8.0%; Score 491.2; DB 9; Length 545;
                  Best Local Similarity 93.9%; Pred. No. 6.2e-105;
                  Matches 511; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
QY 250 CCCAAGAGAGAGAGAGAGAGTATCGTGAACAGAGAAATTAATATATAGAGAACTT 309
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2 CCCAAGAGAGAGAGAGAGTATCGTGAACAGAGAAATTAATATATAGAGAACTT 61
QY 310 GCAGAGTTATTTTTCGCAATTTTAAATGATATAGACAACTTAACTTCAAACTGACAA 369
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 62 GCAGAGCTATTTTTCGCAATTTTAAATGATATAGACAACTTAACTTCAAACTGACAA 121
QY 370 TGTGCAATCTTAAAGAACTGTGAAGCAATTCGTACATCAAGACAAAGAGAGAGGA 429
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 122 TGTGCAATCTTAAAGAACTGTGAAGCAATTCGTACATCAAGACAAAGAGAGGA 181
QY 430 GCAGCTGCCACATAGATGAAGTGCAGAGTCAATCTCTACAGGCGAGGCTGTC 489
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 182 GCAGCTGCCACATAGATGAAGTGCAGAGTCAATCTCTACAGGCGAGGCTGTC 241
QY 490 ATCGACAGAGATGCGCTGGGCGCTATGATGCTTGAAGCCCTTGATGGGTTCTTTTGA 549
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 242 ATCGACAGAGAGCGCTGGGCGCTATGATGCTTGAAGCCCTTGATGGGTTCTTTTGA 301
QY 550 GTGAACCTGGAGGCAAGCTGTGTGTGTGTGTCAGAGAAATGTGACACAGATCTAAGTAT 609
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 302 GTGAACCTGGAGGCAAGCTGTGTGTGTGTGTCAGAGAAATGTGACACAGATCTAAGTAT 361

```

```

QY 610 AACCAAGAGAGCTGATGAAACAAAGTATATATAGCATCTTCATGTGGGAGCACACG 669
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 362 AACCAAGAGAGAGCTGATGAAACAAAGCGTATATAGCATCTTCATGTGGGAGCACAC 421
QY 670 GAATTTGTCAAAACCTGCTGCCAAAGTCTATATAGTAATGAGGAGATCTTGCTGGCAA 729
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 422 GAATTTGTCAAAACCTGCTGCCAAAGTCTATATAGTAATGAGGAGATCTTGCTGGCAA 481
QY 730 CCTCCAGGCGGAGACAGCCATCTCAATGTGGATGCTGTAAACCTTACTGAT 789
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 482 CCTCCAGGCGGAGACAGCCATCTCAATGTGGATGCTGTAAACCTGCTGAT 541
QY 790 TCAG 793
    |||
DB 542 CCAG 545

RESULT 16
BE031427          543 bp  mRNA  linear  EST 09-JUL-2000
LOCUS             BE031427
DEFINITION        129992 MARC 1PIG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION         BE031427
VERSION           BE031427.1 GI:8326436
KEYWORDS          EST.
SOURCE            pig.
ORGANISM          Sus scrofa
                  Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
                  Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE         1 (bases 1 to 543)
                  Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
                  Stone,R.T., Heaton,M.P., Grose,W.M., Bennett,G.A., Laegreid,W.W.,
                  and Keele,J.W.
                  Design and use of two pooled tissue normalized cDNA libraries for
                  EST discovery in swine
JOURNAL           Unpublished (2000)
COMMENT           Contact: Smith TPL
                  USDA, ARS, US Meat Animal Research Center
                  PO Box 166, Clay Center, NE 68933-0166, USA
                  Tel: 402 762 4366
                  Fax: 402 762 4390
                  Email: smithemail.marc.usda.gov
                  Single pass sequencing. Bases called and alt-trimmed with phred
                  v0.980904.e. Vector identified by cross_match with the -minscore 18
                  and -mismatch 12 options.
PCR Primers      FORWARD: AGGAACAGCTATGACCAT
                  BACKWARD: GTTTCCCACTCAGCAGC
                  Plate: 66 row: L column: 9
                  Seq primer: ATTTAGTGACACTATAG.
                  Location/Qualifiers
FEATURES          source
                  1..543
                  /organism="Sus scrofa"
                  /db_xref="taxon:9823"
                  /clone_lib="MARC 1PIG"
                  /tissue_type="pooled"
                  /lab_host="DH10B"
                  /note="Vector: pCMV SPORT6; Site.1: XbaI; Site.2: XhoI;
                  library made from pooled tissue from day 11, 13, 15, 20,
                  and 30 embryos."
BASE COUNT       168 a 110 c 140 g 125 t
ORIGIN
Query Match      7.9%; Score 487.6; DB 9; Length 543;
                  Best Local Similarity 93.7%; Pred. No. 4.4e-104;
                  Matches 508; Conservative 0; Mismatches 34; Indels 0; Gaps 0;
QY 250 CCCAAGAGAGAGAGAGAGTATCGTGAACAGAGAAATTAATATATAGAGAACTT 309
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2 CCCAAGAGAGAGAGAGAGTATCGTGAACAGAGAAATTAATATATAGAGAACTT 61
QY 310 GCAGAGTTGATTTTTCGCAATTTTAAATGATATAGACAACTTAACTGACAACTGACAA 369

```



Db	62	GCAGAGCTGATTTTTTGGCAATTTTAAACGATTATAGACAACTTTAACTTCAAACTGACAAA	121
OY	370	TGTGCATCTTTAAAGAAAATGTGTGAAGCAAAATTCAGATCAAAAGAACAGAAAAGCA	429
Db	122	TGTGCAATCTTAAAGAAAATGTGAAGCAAAATTCAGATCAAAAGAACAGAAAAGCA	181
OY	430	GCAGTGTGCAACATAGATGAAGTGCAGAAAGTACAGTATCCTCTACAGGCGAGGGTCTC	489
Db	182	GCAGTGTGCAACATAGATGAAGTGCAGAAAGTACAGTATCCTCTACAGGCGAGGGGCTC	241
OY	490	ATCGACAAGAGATGCGCTGGGGCCCTATGATGCTTGAGGCCCTGATGGGCTTCCTCTTGTGA	549
Db	242	ATCGACAAGAGAGCGCCCTGGGGCCCTATGATGCTTGAGGCCCTGATGGGCTTCCTCTTGTGA	301
OY	550	GTGAACCTGTGAAGGCAACGTTGTGTTTGTGTGACAGAAATGTACACAGTATCTAAGTAT	609
Db	302	GTGAACCTGTGAAGGCAACGTTGTGTTTGTGTGACAGAAATGTACACAGTATCTAAGTAT	361
OY	610	AACCAAGAAAGAGCTATATACAAAGTATATATACATCTTCCATGTTGGGACACACACG	669
Db	362	AACCAAGAAAGAGCTATATACAAAGTATATATACATCTTCCATGTTGGGACACACACG	421
OY	670	GAATTTGTCAAAAAAAGCTGTGCGCAAGTCTATGATAAATGGGGGATCTTGCTGGCGAA	729
Db	422	GAATTTGTCAAAAAAAGCTGTGCGCAAGTCTATGATAAAGGGGATCTTGCTGGCTGAA	481
OY	730	CTCTCCGAGCGGAACAGCCATACCTTCAATTCGGAATGCTGTAAACCTTTTACCCTGAT	789
Db	482	CTCTCCGAGCGGAGAGCACACCTTCAATTCGGAATGCTGTAAACCTTCGCTGAT	541
OY	790	TC 791	
Db	542	CC 543	

RESULT	17
LOCUS	BB628347
DEFINITION	BB628347 RIKEN full-length enriched, adult male urinary bladder Mus
ACCESSION	musculus cdna clone 9530095N19.5, mRNA sequence.
VERSION	BB628347
KEYWORDS	BB628347.1 GI:16465803
SOURCE	EST.
ORGANISM	house mouse.
REFERENCE	Mus musculus
AUTHORS	Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi: Mammalia: Eutheria: Rodentia: Scuriognathi: Muridae: Murinae: Mus: 1 (bases 1 to 655)
TITLE	Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Horii,F., Ishii,Y., Ito,M., Kawai,J., Kono,H., Kouda,M., Koyu,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,T., Shibata,K., Shingawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
JOURNAL	RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
COMMENT	Unpublished (2001) Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/ Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000) wagii,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,

Query Match	Best Local Similarity	Matches	Conservative	Mismatches	Indels	Gaps
3032	AGAGTTCCGCTGTGAGACTACCTGTGCTGTACCCACGAGTGCATGGAACGGCCAGTCC	3091				
Db	1	AGAGTCACGCTGTGAGACTACTTGTGCTGTACACACGTGTGCTGTGAACGACAGTCC	60			
3092	AAGGAGTATGATTTGGAAACCCAGCAGCCAGCATCCCATGAGGCCAGCAGCCAGCTG	3151				
Db	61	AAGGAGCATATTCGGAACCCAAAGCAGCATCCCATGAGGCCAGCAGCCAGCTG	120			
3152	GCCAAAGCAGACGGCTGTAGTCTCAGGTCAGGTCATGAATATGAGGCGCATCTGAATTTAGAGATGA	3211				
Db	121	GCCAAAGCAGATGCTTATAGTCTCAGGTCATGAATATGAGGCGCATCTTCTAGTTAGAGATGA	180			
3212	ACATGGGGGGAACCTCAGTATAGCCAAACCAAGCTCTCTCAAAATCAGATCCCATGGC	3271				
Db	181	ACATGGGAGAACCTCAGTATATATCAACAGCAGGCCCTCGAACCACAACTGCCCGCTGC	240			
3272	CTGAAGCATCTGCTCTATAGACCAAGCGCTTTTGGCAGCCAAACAGGCAAGCCATTTG	3331				

Db 241 CTGAGAGCATCTGCTATATAGACAGGAGCATGTTGGCAGCAGAAACAGGAGCCCTTGG 300  
Qy 3332 GCAGTTCTCCAGATGATGCTATATGTCACATCTCTGACGTAGTCTCGAGATGATGAG 3391  
Db 301 GAGAGTCCCTTATGACCTGCTGTGTGTCACATCTCTGACAGAGTGCACAGAGATGAG 360  
Qy 3392 GAGCTCTCTGAGCAGCTGTATCTGGCCTTGGGAAATTTGATGAGCCCTGAGAGATGG 3451  
Db 361 GGGCTCTCTTACAGAGTGTATCTGCTGGGAAATTTGATGAGCCCTTGGAGAGATGG 420  
Qy 3452 ATAGAGCCTTAGAATACCCGAACTGTGACGACAGGAGCCAGAGATGATGAGAGCACT 3511  
Db 421 ATAGAGCCTTAGGAGATGACGAGATGTGTGACGAGAGCACTGTGATGAGAGCACT 480  
Qy 3512 TCTCAAGTACAGATTCACATCATCTGAGCAGAGAGGCGCCCTTTTCCACAGCACT 3571  
Db 481 TCTCAAGTACAGAGTCCACATCATCTGAGCAGAGAGGCGCCCTTTTCCACAGCACT 540  
Qy 3572 ATGCATCTCAGCACAATAATGGCCAGGTAGCTATTCTCCATGCAAG-ATCCAAACTTT 3630  
Db 541 ACCCATTTTAAGCACAATAATGGCGAGGTGCTATATCCCTGCAAGATCCAAACTTT 600  
Qy 3631 CACACCATGGGACAGCGGCTTA 3652  
Db 601 TACACCATGGGAAAGCGGCCA 622

RESULT 18  
Bg835556 559 bp mRNA linear EST 25-MAY-2001  
LOCUS Bg835556 354744 MARC 1Pig Sus scrofa cDNA 5', mRNA sequence.  
DEFINITION Bg835556  
ACCESSION Bg835556.1 GI:14201722  
VERSION EST.  
KEYWORDS  
SOURCE pig.  
ORGANISM Sus scrofa  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.  
REFERENCE 1 (bases 1 to 559)  
AUTHORS Fahrenkrug,S.C., Freiling,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,  
Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Jaegerfeld,W.W.  
and Keele,J.W.  
Design and use of two pooled tissue normalized cDNA libraries for  
EST discovery in swine  
Unpublished (2000)  
CONTACT: Smith TPL  
USDA, ARS, US Meat Animal Research Center  
PO Box 166, Clay Center, NE 68933-0166, USA  
Tel: 402 762 4366  
Fax: 402 762 4390  
Email: smith@email.marc.usda.gov  
Single pass sequencing. Bases called and alt trimmed with phred  
v0.980904.e. Vector identified by cross\_match with the -minscore 18  
and -mismatch 12 options.  
PCR Primers  
FORWARD: AGGAACACAGCTATGACCAAT  
BACKWARD: GTTTCACAGTCAGCAGC  
Plate: 117 row: D column: 21  
Seq primer: ATTAGGTGACACTAAG.  
Location/Qualifiers  
1..559  
/organism="Sus scrofa"  
/db\_xref="taxon:9823"  
/clone\_lib="MARC 1Pig"  
/tissue\_type="pooled"  
/lab\_host="DH10B"  
/note="Vector: pCMV SPORT6; Site\_1: XbaI; Site\_2: XhoI;  
Library made from pooled tissue from day 11, 13, 15, 20,  
and 30 embryos."

BASE COUNT 171 a 139 c 149 g 100 t  
ORIGIN

Query Match 7.8%; Score 483; DB 10; Length 559;  
Best Local Similarity 92.7%; Pred. No. 5,4e-103;  
Matches 518; Conservative 0; Mismatches 40; Indels 1; Gaps 1;  
Qy 2170 GGTACTTGCCCTGGTCTGGGCTACATGGAACCTGCTCA-GGAGAACATTAAT 2228  
Db 1 GGCACCTGCTGGGCTCTGGGGGACACAGGAGACCTGCTCAAGGAAACATTAAGT 60  
Qy 2229 TTGGCAGACACTCTTGGCAGGAGAGAGTCCCTGTGACTTGGCCAAATTAACAGAGA 2288  
Db 61 TTGGCAGACACTGTGGCAGGATAGCAGCTCTCTGTGGACTTGGGAAATTAACCTGCGGA 120  
Qy 2289 ACCCAGGCAAGACCTGAGCCAGAGTCCAGCAGCACAGTCTCTGATCGATCGAAGTAC 2348  
Db 121 ACCCAGGCAAGAGCTGTAACCCAGAGGCCAGCAGCAGCCCTGTGATCCGAGGGAG 180  
Qy 2349 TATTAAAGAGAGCGGTGAGGCCCAAGAGAAAGAAATGCACTACTTGGCTATTTGCT 2408  
Db 181 GATTAAAGAGAGCCAGGTAGGCCCAAGAGAAAGAAATGCACTACTTGGCTATTTGCT 240  
Qy 2409 AGATAAGATGATTAAGATATTTGTTTACCAGAAATTAACCCCAACTTGAAGACT 2468  
Db 241 CGATAAGACATTAAGATATGCTTTACCGGAAATTAACCCCAACTTGAAGACT 300  
Qy 2469 GGCAGTAAGACAGATCTGCCAGTAAACAATAATTAATGCAATGAATACTGAGAGGA 2528  
Db 301 GGCAGTAAGACAGACCTGCCAGTAAACAATAATTAATGCAATGAATACTGAGAGGA 360  
Qy 2529 GGAGATGAGCTTGGAGCCGTGGAGCAGCTGGAGTGGAGCACTTGGAGAGAT 2588  
Db 361 GGAGATGAGCTTGGAGCCGTGGAGCAGCTGGAGTGGAGCACTTGGAGAGAT 420  
Qy 2589 TTGGATGATTTGGCAGAAATAGTCAATTAACAGAGCTTTTCCAGACAGAGGCCAGGCC 2648  
Db 421 TTGGATGATTTGGCAGAAATAGTCAATTAACAGAGCTTTTCCAGACAGAGGCCAGGCC 480  
Qy 2649 CCTGCTGATGATGTTGACAGAGCCATCATTAATGCTATGATGCAACTGACAGCTGA 2708  
Db 481 CCTGCTGATGATGTTGACAGAGCCATCATTAATGCTATGCAACTGACAGCTGA 540  
Qy 2709 AAGACGCGCTGCACACT 2727  
Db 541 AAGACGCGCGTTACACT 559

RESULT 19  
AM465013 542 bp mRNA linear EST 24-FEB-2000  
LOCUS AM465013/BP230017B10B11 Soares normalized bovine placenta Bos taurus cDNA  
DEFINITION BP230017B10B11 5', mRNA sequence.  
ACCESSION AM465013  
VERSION AM465013.1 GI:7035181  
KEYWORDS EST.  
SOURCE cow.  
ORGANISM Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovidae; Bovinae; Bos.  
REFERENCE 1 (bases 1 to 542)  
AUTHORS Lewin,H.A., Soares,M.B., Rebeiz,M., Pardinas,J., Liu,L. and Larson,  
J.H.  
TITLE Bovine ESTs  
JOURNAL Unpublished (2000)  
CONTACT: Lewin, H. A.  
W. M. Keck Center for Comparative and Functional Genomics  
University of Illinois at Urbana-Champaign  
340 Edward R. Madigan Laboratory, 1201 W. Gregory Dr., Urbana, IL  
61801, USA  
Tel: 217 333 5998  
Fax: 217 244 5617  
Email: h-lewin@uiuc.edu  
Funding for cattle EST sequencing was provided by the USDA National  
Research Initiative, Animal Genome Resource Grant AG 99-3205-8534

to H. A. Lewin and J. E. Womack. Base Calling/Quality Scores: PHRED from Washington University Genome Center. Vector Trimmi g: Cross-match from Washington University Genome Center PHRAP suite. Sequences submitted are vector free and at least 200 bp in length.

PCR Primers  
 FORWARD: TAATACGACTCCTATAGG  
 BACKWARD: ATTAACCTCTACAAAG  
 Insert Length: 542 Std Error: 0.00  
 Plate: Bp30017B10 row: B column: 11  
 Seq primer: AGCGATACAAATTCACACAGA  
 High quality sequence stop: 542.

## FEATURES

source

1. 542  
 Location/Qualifiers  
 /organism="Bos taurus"  
 /db\_xref="taxon:9913"  
 /clone="BP230017B10B11"  
 /clone\_lib="Soares normalized bovine placenta"  
 /sex="female"  
 /lab\_host="DH10B"  
 /note="Organ: placenta; Vector: pT73Pac; Site: 1: EcoRI; Site\_2: NotI; The cDNA library was contributed by the Soares laboratory and it was constructed and normalized as described by Bonaldo, M.F., Lennon, G. and Soares, M.B. (1996), Genome Research 6(9): 791-806."  
 BASE COUNT 115 a 141 c 111 g 173 t 2 others  
 ORIGIN

Query Match 7.8%; Score 480.8; DB 9; Length 542;  
 Best Local Similarity 92.8%; Pred. No. 1.8e-102;  
 Matches 503; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

OY 124 TGACCTGTTTACGAGCGATGTCGATGTCACAGACGATGGGAGAAAT 183  
 Db 542 TGACCTGTCAGTCACAGTCGATGTCGATGTCACAGACGATGGGAGAAAC 483  
 OY 184 ACTCTGACCCCTCCAGGAGACAGCAAGAAAGCGAAGATGCTGACCACTTGA 243  
 Db 482 ACCCTGACCCATCCAGGAGAGACGGGAAAGCCCAAGAAATGCTGACCACTTGA 423  
 OY 244 CCCAGCCCCAAAGAGACACTGAAACGTAATCGTGAACAGGAATTAATATTAAGA 303  
 Db 422 CCCAGCCCCAAAGAGAGACCGAGAAACGTAACCGTGAACAGGAATTAATATTAAGA 363  
 OY 304 GAACCTGACAGCTGATTTTTCGAAATTTTATGATATGACAACTTAACTCAACCT 363  
 Db 362 GAACCTGACAGCTGATTTTTCGAAATTTTATGATATGACAACTTAACTCAACCT 303  
 OY 364 GACAAATGTCAACTTAAAGAACTGTAAGCAAAATTCGACATCAAGAACAGAG 423  
 Db 302 GACAAATGTCAACTTAAAGAACTGTAAGCAAAATTCGACATCAAGAACAGAG 243  
 OY 424 AAAGCAGACGCTCCACATAGATGAGAGAGAGATGATGATGCTTACAGGAG 483  
 Db 242 AAAGCAGACGCTCCACATAGATGAGAGAGAGATGATGATGCTTACAGGAG 183  
 OY 484 GGTGTCATGACAGAGAGCGCTGGGGCTATGATGCTTGAAGCCCTGATGGGCTTTC 543  
 Db 182 GGTGTCATGACAGAGAGCGCTGGGGCTATGATGCTTGAAGCCCTGATGGGCTTTC 123  
 OY 544 TTTGTAGTGAACCTGGAAGGCAACGTTGTGTTGTGTCAGAAATGTCACAGTATCTA 603  
 Db 122 TTTGTAGTGAACCTGGAAGGCAACGTTGTGTTGTGTCAGAAATGTCACAGTATCTA 63  
 OY 604 AGGTATACCAAGAGACTGATGAACAAAGTGTATATAGCATTTGATGTTGGGAGC 663  
 Db 62 AGGTATACCAAGAGACTGATGAACAAAGTGTATATAGCATTTGATGTTGGGAGC 3  
 OY 664 CA 665  
 Db 2 CA 1

## RESULT 20

LOCUS AM298513

DEFINITION UT-H-BW0-ajm-d-02-0-UT-s1 NCI\_CGAP\_Sub6 Homo sapiens cDNA clone

ACCESSION AM298513

VERSION AM298513.1

KEYWORDS GI:6705149

SOURCE EST.

ORGANISM human.

REFERENCE 1 (bases 1 to 484)

AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

TITLE NCI\_CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

COMMENT Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Oligo-dT track not found. Not I site shown in beginning of sequence

is likely internal to the message. cDNA library preparation: M.B.

Soares Lab Clone distribution: NCI\_CGAP clone distribution

Information can be found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/biopr/image/image.html

Seq primer: M13 forward

POLYA-No.

## FEATURES

source

1. 484  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2732210"  
 /clone\_lib="NCI\_CGAP\_Sub6"  
 /lab\_host="DH10B (Life Technologies)"  
 /note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; NCI\_CGAP\_Sub6 is a subtracted library derived from BW, which consists of a mixture of four normalized libraries: NCI\_CGAP\_Br50, NCI\_CGAP\_Lu13, NCI\_CGAP\_Ov18, GRC1. The NCI\_CGAP\_Sub6 library had 7 million recombinants. A single-stranded DNA preparation of BW was used as a tracer in a subtractive hybridization with a driver comprising: the IMAGE pool (NCI\_CGAP\_Kid3 pool 1 LHAM 3334-3337, 3682-3683, 3798-3803 (IMAGE Clonoids 1322376-1323911, 1456008-1456775, 1500552-1502855); NCI\_CGAP\_Kid5 pool 1 LHAM 3338-3342, 3722-3725, 3776-3778 (IMAGE Clonoids 1323912-1325831, 1471368-1472903, 1492104-1493255); NCI\_CGAP\_Lu5 pool 1 LHAM 3575-3582, 3851-3854 (IMAGE Clonoids 1414920-1417991, 1520904-1522439); NCI\_CGAP\_GC4 pool 1 LHAM 3164-3167, 3716-3720, 3733-3735 (IMAGE Clonoids 1257096-1258631, 1469064-1470983, 1475592-1476743); NCI\_CGAP\_Pr22 pool 1 LHAM 2457-2459, 2758-2759, 3062-3068 (IMAGE Clonoids 985608-986759, 1101192-1101959, 1217928-1220615); NCI\_CGAP\_Co10 pool 1 LHAM 2644-2653, 2871-2872 (IMAGE Clonoids 1057416-1061255, 1144584-1145351). (50% of the driver population), plus a pool of 3,840 arrayed clones from NCI\_CGAP\_Sub1 (IMAGE Clonoids 2708616-2710535) and NCI\_CGAP\_Sub2 (IMAGE Clonoids 2710536-2712455) (20% of the driver population), plus a pool of 11,136 clones from NCI\_CGAP\_Sub3 (IMAGE Clonoids 2712456-2723591) (30% of the driver population). Subtraction was performed as previously described [Bonaldo, Lennon & Soares (1996): Normalization and Subtraction: Two Approaches To Facilitate Gene Discovery. Genome Research 6, 791-806.  
 TAG\_LIB=NCI\_CGAP-Lu13  
 TAG\_TISSUE=Lung  
 TAG\_SEQ=GCCGC"  
 BASE COUNT 156 a 109 c 121 g 98 t  
 ORIGIN

Query Match 7.8%; Score 480; DB 9; Length 484;  
 Best Local Similarity 100.0%; Pred. No. 2.7e-102;

Matches 480; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2126 CGCCCTTACGACCTCTTTGCGGATACAAACAAGCTCCACAGGTAGCTTCGTTG 2185  
|||||  
Db 5 CGCCCTTACGACCTCTTTGCGGATACAAACAAGCTCCACAGGTAGCTTCGTTG 64  
|||||

QY 2186 CTGGGCTTACACATGGAACCTCGCTCAAGAGAAAGCATTAATTTTGCACAGACTCTTCG 2245  
|||||  
Db 65 CTGGGCTTACACATGGAACCTCGCTCAAGAGAAAGCATTAATTTTGCACAGACTCTTCG 124  
|||||

QY 2246 AGGACAGACAGTTCCTCTGTGACCTTGCCCAAGTTAAACAGAGAGACAGGCAAGAC 2305  
|||||  
Db 125 AGGACAGACAGTTCCTCTGTGACCTTGCCCAAGTTAAACAGAGAGACAGGCAAGAC 184  
|||||

QY 2306 TGAAGCAGAGATCCAGACAGACAGCTCTGATCAGAAAGTATTTAAACAAGAGCCG 2365  
|||||  
Db 185 TGAAGCAGAGATCCAGACAGACAGCTCTGATCAGAAAGTATTTAAACAAGAGCCG 244  
|||||

QY 2366 TGAAGCAGAGAGAAAGAGAAATGCACTACTTGGCTATTGCTAGATTAAGATGATCTA 2425  
|||||  
Db 245 TGAAGCAGAGAGAAAGAGAAATGCACTACTTGGCTATTGCTAGATTAAGATGATCTA 304  
|||||

QY 2426 AAGATATGTTGTTACAGAAATTAACCCCAACTTGAGAGCTGAGACATGAACAGATC 2485  
|||||  
Db 305 AAGATATGTTGTTACAGAAATTAACCCCAACTTGAGAGCTGAGACATGAACAGATC 364  
|||||

QY 2486 CTGCCAGTAAACAAATTAATAGCAATGAATACTGAGAGAGAGATGAGCTTTGAGC 2545  
|||||  
Db 365 CTGCCAGTAAACAAATTAATAGCAATGAATACTGAGAGAGAGATGAGCTTTGAGC 424  
|||||

QY 2546 CTGGTGAACAGCTGAGCTGAGCTGAGCAACTTTGGAGAGATTTGGATGATTTGCAGA 2605  
|||||  
Db 425 CTGGTGAACAGCTGAGCTGAGCTGAGCAACTTTGGAGAGATTTGGATGATTTGCAGA 484  
|||||

RESULT 21  
LOCUS BM090425 555 bp mRNA linear EST 19-NOV-2001  
DEFINITION BM090425 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.  
ACCESSION BM090425  
VERSION BM090425.1 GI:17001053  
KEYWORDS EST.  
SOURCE cow.  
ORGANISM Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovidae; Bovinae; Bos.  
1 (bases 1 to 555)  
Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,  
Cassas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett,  
G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,  
Perlea,G., Holt,I., Karameycheva,S., Liang,F., Quackenbush,J. and  
Keefe,J.W.  
Sequence evaluation of four pooled-tissue normalized bovine cDNA  
libraries and construction of a gene index for cattle  
Genome Res. 11 (4), 626-630 (2001)  
21180013  
Contact: Smith TPL  
USDA, ARS, US Meat Animal Research Center  
PO Box 166, Clay Center, NE 68933-0166, USA  
Tel: 402 762 4366  
Fax: 402 762 4390  
Email: smith@email.marc.usda.gov  
Single pass sequencing. Bases called and alt-trimmed with phred  
v0.980904.e. Vector identified by cross\_match with the -mnscore 18  
and -mismatch 12 options.  
PCR primers  
FORWARD: AGGAACACAGCTATGACCAT  
BACKWARD: GTTTCCAGTCACGACG  
Plate: 98 row: F column: 6  
Seq primer: ATTTAGTGACACTATAG.  
Location/Qualifiers  
1. 555

/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/clone\_lib="MARC 3BOV"  
/tissue\_type="pooled"  
/lab\_host="DH10B"  
/note="vector: pCMV SPORT6; site.1: xbaI; site.2: xhoI;  
library made from pooled tissue from marrow, alveolar  
macrophage, ovary, fetal semitendinosus muscle, and fetal  
longissimus muscle."  
BASE COUNT 155 a 165 c 144 g 91 t  
ORIGIN

Query Match 7.8%; Score 478.2; DB 10; Length 555;  
Best Local Similarity 91.4%; Pred. No. 7.3e-102;  
Matches 507; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

QY 2787 AGGCAACCTGGCAGGTTATTTGCGCAACAGAAATTACCACTTGACATACATTCGAAG 2846  
|||||  
Db 1 AGGCAACCTGGCAGGTTATTTGCGCAACAGAAATTTCCTGACATCATTGCAAG 60  
|||||

QY 2847 CCCACTGGTGTGACCTTTCCACCAATCAGAAACAGTATGCTTCACTGATGATACC 2906  
|||||  
Db 61 CCCAAGGGGCGCTGACCTTTCCACCAATCAGAAACAGTATGCTTCACTGATGATACC 120  
|||||

QY 2907 TCAGCAGAGATGATGGTATATCAAGGATGATAGGAAACCAAGAAATTTAGGAACAG 2966  
|||||  
Db 121 TCAGCAGAGATGATGGTATATCAAGGATGATAGGAAACCAAGAAATTTAGGAACAG 180  
|||||

QY 2967 TAGCAGAGATGATGTTGTAACAGTCTTCTGGGCTACTATGCTTGGAGAAATGGCC 3026  
|||||  
Db 181 TAGCAGAGATGATGTTGTAACAGTCTTCTGGGCTACTATGCTTGGAGAAATGGCC 240  
|||||

QY 3027 ACCGACAGTGTGCTGTGAGAGTCACTGTGTGTACACACAGTGCATGAACCGGCC 3086  
|||||  
Db 241 GCCCAGAGTCTGCGGTGAGGTCACTGTGTGTACACAGTGCATGAACCGGCC 300  
|||||

QY 3087 AGTCCAAGAGATGATGATTTGGAACCCAGCAGCAGATCCCATGAGGCCAGCA 3146  
|||||  
Db 301 AGTCCAAGAGATGATGATTTGGAACCCAGCAGCAGATCCCATGAGGCCAGCA 360  
|||||

QY 3147 GCCTGGCCAAAGACAGACGCTTCAGTCAAGTCAATATGAGGCCATGGAATTA 3206  
|||||  
Db 361 GCCTGGCCAAAGACAGACGCTTCAGTCAAGTCAATATGAGGCCATGGAATTA 420  
|||||

QY 3207 GATGAACATGAGGGGGAGCTCAGTATAGCAACCAAGCTCTCCAAATCAGACTGCC 3266  
|||||  
Db 421 GATGAACATGAGGGGGAGCTCAGTATAGCAACCAAGCTCTCCAAATCAGACTGCC 480  
|||||

QY 3267 ATGGCTGAAAGCATCTCTGCTATAGACCAAGGCTTTTGCACGCAAAACAGGAGCC 3326  
|||||  
Db 481 ATGGCTGAAAGCATCTCTGCTATAGACCAAGGCTTTTGCACGCAAAACAGGAGCC 540  
|||||

QY 3327 ATTTGGCAGTTCGC 3341  
|||||  
Db 541 ATTTGGCAGTTCGC 555  
|||||

RESULT 22  
LOCUS BG172538 777 bp mRNA linear EST 06-FEB-2001  
DEFINITION BG172538 60233736F1 NCI\_CGAP\_Mam1 Mus musculus cDNA clone IMAGE:4456760 5',  
mRNA sequence.  
ACCESSION BG172538  
VERSION BG172538.1 GI:12679241  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 777)  
NIH-MGC http://mgc.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
TITLE

JOURNAL  
COMMENT Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: rgs@pds-remail.nih.gov  
Tissue Procurement: Gilbert Smith, Ph.D.  
CDNA Library Preparation: Life Technologies, Inc.  
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNLN at:  
http://image.llnl.gov  
Plate: LNLN0252 row: h column: 09  
High quality sequence stop: 661.  
Location/Qualifiers  
1. 777  
/organism="Mus musculus"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:4456760"  
/clone\_lib="NCI CGAP Mam1"  
/issue\_type="tumor, biopsy sample"  
/dev\_stage="10 months, virgin"  
/lab\_host="DH10B"  
/note="Organ: mammary; Vector: pCMV-SPORT6; Site\_1: Salt;  
Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
Library constructed by Life Technologies. Investigator  
providing samples: Gilbert Smith, NIH"

BASE COUNT 193 a 246 c 199 g 139 t

ORIGIN

Query Match 7.7%; Score 473.6; DB 10; Length 777;  
Best Local Similarity 85.3%; Pred. No. 9.5e-101;  
Matches 62; Conservative 0; Mismatches 94; Indels 13; Gaps 8;

QY 3286 CCTATAGACAGGCGCTTTTGGCCAGCAAGAGCCATTGGCAGTCTCCAGAT 3345  
|||||  
Db 1 CCTATAGACAGGCGCTTTTGGCCAGCAAGAGCCATTGGCAGTCTCCAGAT 60

QY 3346 GACTTGTATGTCCATCCGAGCTGAGTCTCCAGTGTAGAGGAGCTCTCTGAC 3405  
|||||  
Db 61 GACTTGTATGTCCATCCGAGCTGAGTCTCCAGTGTAGAGGAGCTCTCTGAC 120

QY 3406 CAGCTGATCTGCGCTTGGAGATTTTGGATGGCTGAGAGATGATGAGCCTTAGA 3465  
|||||  
Db 121 CAGCTGATCTGCGCTTGGAGATTTTGGATGGCTGAGAGATGATGAGCCTTAGA 180

QY 3466 ATACCCGAACTGTGACAGCCAGAGCAGATGATCCAGAAAGTTCTCAAGTCAGAT 3525  
|||||  
Db 181 ATACCCGAACTGTGACAGCCAGAGCAGATGATCCAGAAAGTTCTCAAGTCAGAT 240

QY 3526 TCCAAATCATGCTGGAGAGCAAGAGCGCGCTTTCCACAGAGATGATGATTCAGCA 3585  
|||||  
Db 241 TCCAAATCATGCTGGAGAGCAAGAGCGCGCTTTCCACAGAGATGATGATTCAGCA 300

QY 3586 CAAATGCGCCAGAGGTACTATCTCCATGCAAGATCCAAATCTTCAACCATGGAGAG 3645  
|||||  
Db 301 CAAATGCGCCAGAGGTACTATCTCCATGCAAGATCCAAATCTTCAACCATGGAGAG 360

QY 3646 CGGCTGATGATGACACTCCGATGACAGCCAGCGGGCGCTCAGGCCAGCGGCTTA 3705  
|||||  
Db 361 CGGCTGATGATGACACTCCGATGACAGCCAGCGGGCGCTCAGGCCAGCGGCTTA 420

QY 3706 GTGCAAGAACAGCAATCAACTTAAGCTTCACTTCAAGCTGCGCTTCAAGCAGCAG 3765  
|||||  
Db 421 GTGCAAGAACAGCAATCAACTTAAGCTTCACTTCAAGCTGCGCTTCAAGCAGCAG 479

QY 3766 AATGCGCAGCCACTTGAATCA -AATGCGAATGTTTCCAAATG -GAATCTGACTGCA 3823  
|||||  
Db 480 AATGCGCAGCCACTTGAATCACTGATCAGATCAGAGTGTCTCAATATGAACTGACTGCA 539

QY 3824 GGGCTGAGTACCAACAGAGC -ACCTATTATGACAGATGCTGGGCCAGAGACAGAG 3881  
|||||  
Db 540 GGGCTGAGTACCAACAGAGCCTTCCATATTATGACAGATGCTGG -CCAGAGGACAGAG 598

QY 3882 GGAATCTGAGACAGACATCTTCGACAGAGCAATGATGACAGCAAGTTCAGCA 3941  
|||||  
Db 599 GGAATCTGAGACAGACATCTTCGACAGAGCAAGTTCAGCAAGTTCAGCA 653

QY 3942 AGCACTTTG-ATGATGAGAGACAGAGGTTGATATGACACCAAGATGCTGCTTA 4000  
|||||  
Db 654 AGCACTTTGATGATGAGAGAGCGAGCGGCTTGAATGATGAC-CCAGTATGATGCTCCCG 712

QY 4001 GTGCTATGCC 4010  
|||||  
Db 713 GTGCTATGCC 722

RESULT 23  
BG990265 592 bp mRNA linear EST 13-JUN-2001  
LOCUS BG990265  
DEFINITION MR2-HT1162-240101-011-H09 HT1162 Homo sapiens CDNA, mRNA sequence.  
ACCESSION BG990265  
VERSION BG990265.1 GI:14394335  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 592)  
AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,  
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,  
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bata, G.S., Simpson, D.H.,  
Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,  
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and  
Simpson, A.J.  
Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
20202663  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=MR2&tl2=MR2-HT1162-  
240101-011-H09&tl3=2001-01-24&tl4=1)  
Seq primer: puc 18 forward  
High quality sequence start: 12  
High quality sequence stop: 300.  
Location/Qualifiers  
1. 592  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="HT1162"  
/dev\_stage="Adult"  
/note="Organ: head\_neck; Vector: puc18; Site\_1: SmaI;  
Site\_2: SmaI; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the puc 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."

BASE COUNT 166 a 154 c 175 g 97 t

ORIGIN

Query Match 7.5%; Score 464.6; DB 10; Length 592;  
Best Local Similarity 92.4%; Pred. No. 1.2e-98;  
Matches 544; Conservative 0; Mismatches 39; Indels 6; Gaps 5;

QY 1789 GGGCAGGGGCTCATATGAGGCTTCCTTGGCTTACACAGACCTAAAAATGGCAATTG 1848

```
Db 4 GGAGACCGGGCTTCATTAGGTCATCGTTGCTTCAACACACCTAAATAAGGGCAATTTC 63
OY 1849 CAAACCTCCCACTTATATGAATCCCTCCCACTACAGAA-GATGGGAAGCTTGACATC 1907
Db 64 CAAACCTCCCACTTATATGAATCCCTCCCACTACAGAAAGGGAGGCTTGACATC 123
OY 1908 AAAAGACTGTTTGGACTATATGGGAGCCCTCTGAAAGTACAACTGGACACAGAGAG 1967
Db 124 AAAAGAGGTTTGGACTATATGGGAGCCCTCTGAAAGTACAACTGGACACAGAGAG 183
OY 1968 CAGCTGCCATCTTGAGAGAGCAAAAAGAACAA-TGACCCCAACTGCCCCGGCGCTGA 2026
Db 184 CAGCTGCCATCTTGAGAGAGCAAAAAGAACAAAGTACCCCAACTGCCCCGGCGCTTA 243
OY 2027 GCAGTGA--GAGAGCTGACGGGAGAGACAGACTGCATGACAGCAAGGGGAGACCAACT 2084
Db 244 GCAGTACAGAGACGCTGACGGGAGAGACAGACTGCATGACAGCAAGGGGAGACCAACT 303
OY 2085 CCTGACAGCTGCTGACCAACCAATCTGATCAGA-TGAGACCCCTGCGCCCTTACAGCTCTT 2143
Db 304 CCGCGCGCTGCTGATCACCACCAATCTGATCAGAGTGAAGCCCTGCGCCCTTACAGCTCTT 363
OY 2144 TGTGGAATACAAACAAAGACTCCACAGGTAGCTTCCCTGCTTGGCTTACACATGAA 2203
Db 364 GGTGCGATACAAACAAAGACTCCACAGGTAGCTGCGCGGCTTGGGCTTACACATGAA 423
OY 2204 CCTGCGTCAAGAGAGCATAAATTTTGCAGAGCTCTTGCAGAGACAGAGTCCCTG 2263
Db 424 CCTGCGTCAAGAGAGCATAAATTTTGCAGAGCTCTTGCAGAGACAGAGTCCCTG 483
OY 2264 TGGACTTGGCCAAATTAAACAGCAAGCCACAGCAAAAGCACTGAGCCAGAGTCCAGCA 2323
Db 484 TGGACTTGGCCAAATTAAACAGCAAGCCACAGCAAAAGCACTGAGCCAGAGTCCAGCA 543
OY 2324 GCACAGCTCCTGATGATCA-GAAGTGAATTAATAACAGAGCCGCTGAGCC 2371
Db 544 GCCCAGGTCTCTGATCAGAAAGTGAATTAATAACAGAGAGGGGTGAGAC 592

RESULT 24
AA744255 467 bp mRNA linear EST 23-JAN-1998
LOCUS AA744255
DEFINITION ny62911.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1282916 3',
mRNA sequence.
ACCESSION AA744255
VERSION AA744255.1 GI:2785005
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 467)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
Ph.D., Gerald Marti, M.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldi, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/BLNt.at:
www.bio.ln1.gov/bbrp/image/image.html
Insert length: 1583 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 391.
Location/Qualifiers
1.467
```

```
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1282916"
/clone.lib="NCI_CGAP_GCB1"
/tissue_type="germinal center B cell"
/lab_host="DH10B"
/note="Vector: pTR3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IgD-),
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
primed with a Not I - 0190(dt) primer
15'-TGTTACCAATGAGAGTGGAGGAGGCGCCCTCATTTTCTTTTCTTTT-3'
]. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pTR3 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldi."
```

BASE COUNT 151 a 106 c 114 g 96 t

ORIGIN

Query Match 7.5%; Score 462.2; DB 9; Length 467;  
Best Local Similarity 99.4%; Pred. No. 4.2e-98;  
Matches 464; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
OY 2229 TTTCACAGACCTGTGAGAGACAGAGCTTCCCTGTGGACTTGCCCAAGTTAACAGACA 2288
Db 1 TTTCACAGACCTGTGAGAGACAGAGCTTCCCTGTGGACTTGCCCAAGTTAACAGACA 60
OY 2289 AGCCACAGGCAAGACCTGAGCCAGAGATCCACAGCAGCTCTGATCAGAAATGAC 2348
Db 61 AGCCACAGGCAATGACCTGAGCCAGAGATCCACAGCAGCTCTGATCAGAAATGAC 120
OY 2349 TATTTAAACAAGACCGGTGAGCCCAAGAAAGAGAAATGCACTTCTTGCTATTGCT 2408
Db 121 TATTTAAACAAGACCGGTGAGCCCAAGAAAGAGAAATGCACTTCTTGCTATTGCT 180
OY 2409 AGATTAAGATGATCTTAAATATGTTTACAGAAATTAACCCCAACTTGAAGACT 2468
Db 181 AGATTAAGATGATCTTAAATATGTTTACAGAAATTAACCCCAACTTGAAGACT 240
OY 2469 GGACAGTAAAGACAGATCTTCCAGTAAACAAAATTAATAGCAATGAAGAACTGAGAAGA 2528
Db 241 GGACAGTAAAGACAGATCTTCCAGTAAACAAAATTAATAGCAATGAAGAACTGAGAAGA 300
OY 2529 GGAGATGAGCTTTGAGCTGTGACCAAGCCTGGCAGTGAAGCTTGGAGAGAT 2588
Db 301 GGAGATGAGCTTTGAGCTGTGACCAAGCCTGGCAGTGAAGCTTGGAGAGAT 360
OY 2589 TTTCGATGATTTGACAGATATGCAATTAACAGAGCTTTTCCAGACAGAGGCGGCG 2648
Db 361 TTTCGATGATTTGACAGATATGCAATTAACAGAGCTTTTCCAGACAGAGGCGGCG 420
OY 2649 CCTGCTGATCAGTTGACAAGCAAGCCATCATATGACCTCATGC 2695
Db 421 CCTGCTGATCAGTTGACAAGCAAGCCATCATATGACCTCATGC 467

RESULT 25
AM504723 466 bp mRNA linear EST 02-MAR-2000
LOCUS AM504723
DEFINITION UI-HF-BNO-alm-d-01-0-UI.r1 NIH_MGC_50 Homo sapiens cDNA clone
IMAGE:3080089 5', mRNA sequence.
ACCESSION AM504723
VERSION AM504723.1 GI:7142390
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 466)
```



QY	3939	GCAACGAACTTTGATGATGAGAGGACAAAGGGTTTGAATATGACACCAAGCATGGTGGCTCC	3998
Db	352	GCAGCGGACACTCTGATGATGATGAGAGACAGGGCTTTGATGTGACCCCAAGCATGGTGGCTCC	411
QY	3999	TAGTGTATGCCAGCAACTATGAGCAACCCCTTGATTTCCCGAGGCAAAATGCACAGCAATT	4058
Db	412	CGCTGGCGCTACACAGAGCATATGAGCAATTCGCCGATATCCCGAGGACAGGCCAATGCCAGCAATG	471
QY	4059	TCCATTTCTCCCAAACTACAGGAATAGTACAGACCTGATTCACAGGCTTTACTGGGGCTAC	4118
Db	472	CCCATTTCTCTCCGAACCTACGGAATAGTACAA-CTGATCTCTGGCTTTACTGGGGCTAC	530
QY	4119	GACTTCCCGAGAGCCCACTTATGTGACCCCGAATGGCACATACAGAGTCCCATGATGCA	4178
Db	531	GA-TCCCGAGATGCTCTTAAATGTCTCCCGGATG-ACATACTCAGAGTCCCATGATGCA	588
QY	4119	ACAGTCTCAGGCGCAACCCAGCCTATACAGGCCCCCTCCGACATGAATGATGGGGCGCAGG	4238
Db	589	GCAGTCTCAGGCGCAACCC-AGCTTACAGGCGCCTCAGCATGAATGATGGGACAGG	646
QY	4239	GAACTATGGGCGGAAACAGCATGTTTTCACAGAGTCCCGACACACATTTGGGCGAGCAAGC	4298
Db	647	GAGCATGGGTGG-AACAGCATGTGTACAGG--AGTCCACACACACTTGGGACAGCAAGC	703
RESULT	27		
LOCUS	AV666871	556 bp	linear
DEFINITION	AV666871 Bos taurus adipocyte cell line Bos taurus cDNA clone		
ACCESSION	AV666871		
VERSION	AV666871.1	GI:9925901	
KEYWORDS	EST.		
SOURCE	CON.		
ORGANISM	Bos taurus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos. 1 (bases 1 to 556) Takasuga,A., Hirotsune,S., Itoh,R., Itohzono,A., Suzuki,H., Aso,H. and Sugimoto,Y. Establishment of a high throughput EST sequencing system using poly(A) tail-removed cDNA libraries and determination of 35,000 bovine ESTs		
TITLE	Nucleic Acids Res. 29 (22), E108 (2001)		
JOURNAL MEDLINE	21570554		
COMMENT	Contact: Yoshikazu Sugimoto Animal Genetics Division Shirakawa Institute of Animal Genetics Ogikura, Nishigo, Nishi-Shirakawa, Fukushima 961-8061, Japan Tel: 81-248-25-5641 Fax: 81-248-25-5725 Email: kazusugi@cocoa.ocn.ne.jp Single pass sequencing. This clone was obtained from a polyA-deleted cDNA library.		
FEATURES	Location/Qualifiers		
SOURCE	1..556 /organism="Bos taurus" /db_xref="taxon:9913" /clone="E1AD021F12" /clone_lib="Bos taurus adipocyte cell line" /cell_type="an adipocyte cell line" /lab_host="DH10B" /note="vector: pZL1, Site_1: SalI, Site_2: NotI, Poly A was deleted from a NotI site"		
BASE COUNT	127 a 186 c 149 g 94 t		
ORIGIN			

Query Match	7.2%;	Score 444.2;	DB 9;	Length 556;
Best Local Similarity	88.4%;	Pred. No. 7.8e-94;		
Matches 482; Conservative	0;	Mismatches 63;	Indels 0;	Gaps 0;

QY	3331	GGAGTTCTCCAGAGAGATTCCTGATGTCACATCTCCAGTGGAGTCTCCAGAGATGAC	3390
Db	12	GGCAGCTCTCCCGAGATCTGCTGTCTCAATCCCGCAGAGATCTCCAGTATGAA	71
QY	3391	GGAGCTCTCTGGACACAGCTTATCTGGCTTGGCGAATTTTGAATGGCTGGAGAGATT	3450
Db	72	GGCGCTCTCCGAGACAGCTTACATGGCTTTGAGGAACCTTGATGGCTGGAGAGATC	131
QY	3451	GATGAGGCTTTAGGAATACCCGAACCTGCTCAGCCAGAGCCAAAGCATGTGATCCAGAACG	3510
Db	132	GACAGAGCCCTGGGGGATACCCGAGCTCTCAGCAGAGACCAGAGTGTGGACCCGAGACG	191
QY	3511	TTTTCAACTCGAGATTCCAATCATGCTGGAGGAGAGAGGGCCGCTTTCCACAGACG	3570
Db	192	TTTCTCACTCGAGATTCAGCTCATGCTTAAGCAGAAAGACCCCGTGTCCCGAGAG	251
QY	3571	TATGCATCTCGAGCAAAATGGCCAGGATAGCTATTCTCCATGCAAGATCCAAACTTT	3630
Db	252	TATGGCTTCAGGGCGCAGATGGCCAGGGCAGCTACAGCCCATTCAGAGATCCAAACTTC	311
QY	3631	CACACCATGGGACAGCGGCTAGTTATGCGCACACTCCGATGACGCCAGACCGGGCCCTC	3690
Db	312	CACACCATGGGACAGCGGCCAGTTAACCCACCTCCGCATGACGCCCGGCTGGCCCTC	371
QY	3691	AGGCCCCAGGGGCTTAGTGCACAAACAGCCCAATCACTAAGACCTTCACTCAGATGCG	3750
Db	372	AGGCCCCAGGGCTTGTGTCAGAGCGGCCCAACGCTGAGAGGCTTCAACTCAGATGCG	431
QY	3751	CTCCAGACAGCAGCAATCGCCAGCACTTATGAATCAATCAGCAATGTTTCCATGTG	3810
Db	432	CTCCAGACAGCAGCAATCGCCAGCACTTATGAATCAATCAGCAATGTTTCAATGTG	491
QY	3811	AACTTGACTGTGAGGCTGAGATACCAACAGAGCACTATTATGACAGATGCTGGCC	3870
Db	492	AACTTGACTGTGAGGCTGAGATACCAACAGAGCACTATTATGACAGCCAGATGCTGGCC	551
QY	3871	CAGAG 3875	
Db	552	CAGAG 556	

RESULT		28
LOCUS	B6658506	
DEFINITION	B6658506 RIKEN full-length enriched, 12 days embryo eyeball mus	
ACCESSION	muscuscul CDNA clone D23047M20_5', mRNA sequence.	
VERSION	B6658506	
KEYWORDS	B6658506.1 GI:16492331	
SOURCE	EST.	
ORGANISM	house mouse. Mus musculus	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus. 1 (bases 1 to 629)	
AUTHORS	Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Horii,F., Ishii,Y., Ito,M., Kawai,U., Konno,H., Kotada, 'M.', Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki, 'D.', Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toyota.T., Muramatsu,M. and Hayashizaki,Y.	
TITLE	RIKEN Mouse ESTs (Arakawa,T., et al. 2001)	
JOURNAL	Unpublished (2001)	
COMMENT	Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Shihoro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan Tel.: 81-45-503-9222 Fax: 81-45-503-9216  Email: genome-res@sc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/ Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh	



241 AACACCTCTGACCCGTC CAGGGCAGAGACCAGAAACGCCAAGGAATGTCCCGACCAAGCTC 300

Query Ma

Query Match	7.1%;	Score 439.4;	DB 10;	Length 673;
Best Local Similarity	86.9%;	Pred. No. 1.1e-92;		

Matches	552;	Conservative	0;	Mismatches	76;	Indels	7;	Gaps	6;
Qy	3950	TCATGATGAGAGACGAAGGTTGATATATGACACCAAGCATGAGTCCCTAGTGTATGC		4009					
Db	1	TCATGATGAGAGAC--AGGCTTGAATGTGACCCCAAGCAGGTGCTCCCGCTG--CTA		57					
Qy	4010	CAGCAACTATGAGCAACCTCGGATTTCCCGAGCAATATGACAGCATTTTCATTTCTC		4069					
Db	58	CCAGCAGCATGAGCAATCCCGGATCCCGAGGCAATGCGCAGCATGCGCCATTTCTC		117					
Qy	4070	CAAACTACGGAATTAATCTGACCAACCTGATCCAGGCTTACTGGGGCTAGACCTCCCA		4129					
Db	118	CGAATCTAGGGAATTAATCTGACCAACCTGATCCAGGCTTACTGGGGCTAGACCTCC		177					
Qy	4130	GCCCACTTATGTCACCCCAATGACATACAGAGATCCCATGATGACAAAGTCTCAG		4189					
Db	178	GTCTCTTATATGTCCTCCCGATGACATATCTCAGATGCTCCATGATGACAGATCTCAG		237					
Qy	4190	CCAAACCCAGCTATACAGGCCCCCTCGACATTAATGATGGGCGGAGGGAACATGGCG		4249					
Db	238	CCAAACCCAGCTATACAGGCCCCCTCGACATTAATGATGGGCGGAGGGAACATGGCG		297					
Qy	4250	GAAGACGATGTTTCCCAAGCATCCCAACACATTTGGGCGGAGGGAACACCAAGCA		4309					
Db	298	GAAGACGATGTTTCCCAAGCATCCCAACACATTTGGGCGGAGGGAACACCAAGCA		357					
Qy	4310	TGTACAGTAAACATGATGATGTCATGCGGACCAACAGAGTGGCATGAGCA		4369					
Db	358	TGTATATGTAACATGATGATGTCATGCGGACCAACAGAGTGGCATGAGCA		417					
Qy	4370	GCATGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG		4428					
Db	418	GCATGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG		477					
Qy	4429	GGGCTCTCTCCATGCTGCTCCCGAGGATTAATGATGCTGCTGAG--GGGAGGCAAC		4487					
Db	478	GGGCTCTCTCCATGCTGCTCCCGAGGATTAATGATGCTGCTGAG--GGGAGGCAAC		536					
Qy	4488	GTTCACCAACAGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATG		4546					
Db	537	GTTCACCAACAGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATG		596					
Qy	4547	AATATTGCTGACACTGCTGAAGCCAGTTCCTT		4581					
Db	597	AATATTGCTGACACTGCTGAAGCCAGTTCCTT		631					
RESULT	30								
LOCUS	BE170504/C	564 bp	mrna	linear	EST 21-JUN-2000				
DEFINITION	QV4-HT0538-170300-142-h09 HT0538		Homo sapiens	CDNA, mRNA	sequence.				
ACCESSION	BE170504								
VERSION	BE170504.1								
KEYWORDS	EST.								
SOURCE	human.								
ORGANISM	Homo sapiens								
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;								
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.								
	1 (bases 1 to 564)								
	Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Brijones, M. R.,								
	Nagai, M. A., da Silva, W. Jr., Zago, M. A., Bordin, S., Costa, F. F.,								
	Goldman, G. H., Carvalho, A. F., Matsukuma, A., Bala, G. S., Simpson, D. H.,								
	Brustein, A., de Oliveira, P. S., Bucher, P., Jongeneel, C. V., O'Hare								
	M. J., Soares, F., Brentani, R. R., Reis, L. F., de Souza, S. J. and								
	Simpson, A. J.								
TITLE	Shotgun sequencing of the human transcriptome with ORF expressed								
JOURNAL	sequence tags								
MEDLINE	Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)								
COMMENT	20202663								
	Contact: Simpson A.J.G.								
	Laboratory of Cancer Genetics								
	Ludwig Institute for Cancer Research								
	Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,								

FEATURES	source	1..564	/organism="Homo sapiens"	
			/db_xref="taxon:9606"	
			/clone_lib="HT0538"	
			/dev_stage="Adult"	
			/note="Organ: head, neck; Vector: puc18; Site: 1: Sma1;	
			Site: 2: Sma1; A mini-library was made by cloning products	
			derived from ORESTES PCR (U.S. Letters Patent application	
			No. 196,716 - Ludwig Institute for Cancer Research)	
			profiles into the puc 18 vector. Reverse transcription of	
			tissue mRNA and cDNA amplification were performed under	
			low stringency conditions."	
BASE COUNT	136 a	140 c	126 g	162 t
ORIGIN				
Query Match	7.1%	Score 437.2;	DB 9;	Length 564;
Best Local Similarity	93.0%;	Pred. No. 3.5e-92;		
Matches	480;	Conservative	0;	Mismatches 33; Indels 3; Gaps 2;
Qy	1034	CCATGAACACGAGCTGGGAGGACCTGTAAGAAGTGTATTCAGAACTTCATGCCACG		1093
Db	549	CCATGAACACGAGCTGGGAGGACCTGTAAGAAGTGTATTCAGAACTTCATGCCACG		490
Qy	1094	ATGAGAGAGATCTGTGCTTATGCTTAAGAGGATCATGATGATGATGATGATGATG		1153
Db	489	ATGAGAGAGATCTGTGCTTATGCTTAAGAGGATCATGATGATGATGATGATGATG		430
Qy	1154	TGGCATTCAGTCAATCTATGCTTTCCTGTCGATGAGGACCTGTCGTCGACAA		1213
Db	429	TGGCATTCAGTCAATCTATGCTTTCCTGTCGATGAGGACCTGTCGTCGACAA		370
Qy	1214	CGAAGAGCAAACTCATCCCTTCACAGACTACTAATGAACCTCAACTGTAATCTTAC		1273
Db	369	CGAAGAGCAAACTCATCCCTTCACAGACTACTAATGAACCTCAACTGTAATCTTAC		310
Qy	1274	ATATGCTTCACAGAGACGAATGTGTGTGATGATGATGATGATGATGATGATGATG		1333
Db	309	ATATGCTTCACAGAGACGAATGTGTGTGATGATGATGATGATGATGATGATGATG		250
Qy	1334	TGGGGAAGCCACTGATTCATTAACAGCCCTCCCATCGAGCCCTGTGAGTG		1393
Db	249	TGGGGAAGCCACTGATTCATTAACAGCCCTCCCATCGAGCCCTGTGAGTG		190
Qy	1394	GGAACCCAGTTCAGAGATGACCTCAGTATGATTAATTTTCCATTAATGGCCAA		1453
Db	189	GCGTTCACAGTTCAGAGATGACCTCAGTATGATTAATTTTCCATTAATGGCCAA		130
Qy	1454	AGGACAAATGGGATGCCCA--TGGGAGGTTTGGTGGTCT--GGGGAATGAACCATG		1510
Db	129	AGGACAAATGGGATGCCCA--TGGGAGGTTTGGTGGTCT--GGGGAATGAACCATG		70
Qy	1511	TGTACGATGACAGCAACCACTCTCAGAGGTAGTA		1546
Db	69	TGTACGATGACAGCAACCACTCTCAGAGGTAGTA		34
RESULT	31			
LOCUS	BB665333	688 bp	mrna	linear
DEFINITION	BB665333			EST 31-AUG-2001
				Mus

















**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 12, 2002, 03:44:51 : Search time 173.65 Seconds  
(without alignments)  
8707.860 Million cell updates/sec

Title: US-09-842-256-1

Perfect score: 6156  
Sequence: 1 GCGGCGCCGACGCTCGGCTA.....CCACTTTATGGTTGTTTTT 6156

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/2/1na/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/2/1na/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/2/1na/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/2/1na/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/2/1na/PCITUS\_COMB.seq:\*  
6: /cgn2\_6/ptodata/2/1na/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6156	100.0	6156	4 US-08-891-640-1	Sequence 1, Appl1
2	6154.4	100.0	6156	4 US-09-723-535-3	Sequence 3, Appl1
3	363.2	5.9	4547	4 US-09-723-534-3	Sequence 3, Appl1
4	58.6	1.0	2082	2 US-08-785-310A-2	Sequence 2, Appl1
5	55.4	0.9	2078	2 US-08-785-310A-1	Sequence 1, Appl1
6	55	0.9	3545	3 US-08-885-291-54	Sequence 54, Appl1
7	55	0.9	3545	3 US-09-496-672-54	Sequence 54, Appl1
8	55	0.9	5715	4 US-09-107-847-1	Sequence 1, Appl1
9	53.8	0.9	7218	1 US-08-232-463-14	Sequence 14, Appl1
10	51	0.8	7218	1 US-08-232-463-14	Sequence 14, Appl1
11	50.2	0.8	4184	2 US-08-785-310A-4	Sequence 4, Appl1
12	48.6	0.8	7498	2 US-08-816-693A-1	Sequence 1, Appl1
13	48.6	0.8	7498	3 US-08-885-291-1	Sequence 1, Appl1
14	48.6	0.8	7498	4 US-09-496-672-1	Sequence 1, Appl1
15	48.2	0.8	2816	4 US-08-785-241-1	Sequence 1, Appl1
16	47	0.8	4010	2 US-08-785-310A-3	Sequence 3, Appl1
17	45.2	0.7	665	2 US-08-883-795A-36	Sequence 36, Appl1
18	43.4	0.7	3031	1 US-08-785-241-2	Sequence 2, Appl1
19	42.6	0.7	1257	3 US-08-640-906-1	Sequence 1, Appl1
20	42.6	0.7	1257	4 US-09-395-936-1	Sequence 1, Appl1
21	42.6	0.7	19124	2 US-08-487-826B-13	Sequence 13, Appl1
22	42.4	0.7	289	4 US-09-007-005-17	Sequence 17, Appl1
23	42.4	0.7	289	4 US-09-244-796-17	Sequence 17, Appl1
24	41.2	0.7	6243	2 US-09-056-075-1	Sequence 1, Appl1
25	41	0.7	1511	1 US-07-991-867B-8	Sequence 8, Appl1
26	41	0.7	1511	1 US-08-107-755A-8	Sequence 8, Appl1
27	41	0.7	1511	2 US-08-544-332-8	Sequence 8, Appl1

C	28	41	0.7	19124	2	US-08-487-826B-13	Sequence 13, Appl
C	29	40.8	0.7	2989	3	US-09-068-195-10	Sequence 10, Appl
C	30	40.8	0.7	5565	3	US-09-068-195-21	Sequence 21, Appl
C	31	40.2	0.7	15225	2	US-08-892-403A-2	Sequence 2, Appl
C	32	40	0.6	5261	1	US-08-945-806-3	Sequence 3, Appl
C	33	40	0.6	5261	1	US-08-366-051B-3	Sequence 3, Appl
C	34	39.8	0.6	1000	1	US-08-117-083-63	Sequence 63, Appl
C	35	39.4	0.6	615	4	US-08-998-416-186	Sequence 186, Appl
C	36	39.4	0.6	654	5	PCT-US95-06406A-11	Sequence 11, Appl
C	37	39	0.6	6152	4	US-08-973-462-1	Sequence 1, Appl
C	38	38.6	0.6	1023	1	US-08-256-261-16	Sequence 16, Appl
C	39	38.6	0.6	1023	3	US-08-852-299-16	Sequence 16, Appl
C	40	38.2	0.6	4673	1	US-07-638-431-1	Sequence 1, Appl
C	41	38.2	0.6	4673	5	PCT-US92-00018-1	Sequence 1, Appl
C	42	38	0.6	701	4	US-08-998-416-701	Sequence 701, Appl
C	43	38	0.6	246240	4	US-08-724-394A-20	Sequence 20, Appl
C	44	38	0.6	246240	2	US-08-724-394A-21	Sequence 21, Appl
C	45	38	0.6	246240	2	US-08-724-394A-22	Sequence 22, Appl

#### ALIGNMENTS

RESULT 1  
US-08-891-640-1  
: Sequence 1, Application US/08891640  
: Patent No. 6268173  
: GENERAL INFORMATION:  
: APPLICANT: Chambon, Pierre  
: APPLICANT: Gronemeyer, Hinrich  
: APPLICANT: Voegel, Johannes  
: APPLICANT: Lutz, Yves  
: TITLE OF INVENTION: Transcriptional Intermediary Factor-2  
: NUMBER OF SEQUENCES: 14  
: CORRESPONDENCE ADDRESSES:  
: ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.  
: STREET: 1100 New York Avenue, NW, Suite 600  
: CITY: Washington  
: STATE: DC  
: COUNTRY: USA  
: ZIP: 20005-3934  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: PatentIn Release #1.0, Version #1.30  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/891,640  
: FILING DATE: Herewith  
: CLASSIFICATION: 435  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 60/021,247  
: FILING DATE: 12-JUL-1996  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Steffe, Eric K.  
: REGISTRATION NUMBER: 36,688  
: REFERENCE/DOCKET NUMBER: 1383.0130001/EKS  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: 202-371-2600  
: TELEFAX: 202-371-2540  
: INFORMATION FOR SEQ ID NO: 1:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 6156 base pairs  
: TYPE: nucleic acid  
: STRANDEDNESS: double  
: TOPOLOGY: linear  
: MOLECULE TYPE: cDNA  
: FEATURE:  
: NAME/KEY: CDS  
: LOCATION: 163..4554  
: US-08-891-640-1

Query Match	100.0%	Score 6156	DB 4	Length 6156
Best Local Similarity	100.0%	Pred. No. 0		
Matches 6156	Conservative	0	Mismatches 0	Indels 0
QY	1	GGCGGCGGAGCTCGGGCTACAGCTTCGGCGGGCGAAGGTCAGCGCGGACGGCGCGCA	60	
Db	1	GGCGGCGGAGCTCGGGCTACAGCTTCGGCGGGCGAAGGTCAGCGCGGACGGCGCGCA	60	
QY	61	CCTGACGGCGTGACGAGCGACCGCGGATTTTCCTTGAGTTTGGCTACACACTTATAGATC	120	
Db	61	CCTGACGGCGTGACGAGCGACCGCGGATTTTCCTTGAGTTTGGCTACACACTTATAGATC	120	
QY	121	TTCTGCACCTGTTTACAGGACAGCTTGCTATATGTGTTCAGATGAGTGGATGGAGAA	180	
Db	121	TTCTGCACCTGTTTACAGGACAGCTTGCTATATGTGTTCAGATGAGTGGATGGAGAA	180	
QY	181	AATFACCTCGACCCCTCCAGGGCAGAGCAAGAAAGCGGAAGAAATGTCCTGACCAACT	240	
Db	181	AATFACCTCGACCCCTCCAGGGCAGAGCAAGAAAGCGGAAGAAATGTCCTGACCAACT	240	
QY	241	GGACCCAGCGCCCAAAAAGAAACACTGAAAAACGTAACTGTAACAGAGAAATTAATATAT	300	
Db	241	GGACCCAGCGCCCAAAAAGAAACACTGAAAAACGTAACTGTAACAGAGAAATTAATATAT	300	
QY	301	GAGAAACTTGCAGAGTTGATTTTTGCAAAATTTTAAATGATATAGACAACTTAACTTCAAA	360	
Db	301	GAGAAACTTGCAGAGTTGATTTTTGCAAAATTTTAAATGATATAGACAACTTAACTTCAAA	360	
QY	361	CCTGCACAAATGTCACATCTTTAAAGAAACTGTGGAAGCAAAATTCGTACATTCAAAGACAA	420	
Db	361	CCTGCACAAATGTCACATCTTTAAAGAAACTGTGGAAGCAAAATTCGTACATTCAAAGACAA	420	
QY	421	GAGAAAGCAGCAGCTGCCCAACATAGATGAAGTGACAGAACTAGATATATCTCTACAGGG	480	
Db	421	GAGAAAGCAGCAGCTGCCCAACATAGATGAAGTGACAGAACTAGATATATCTCTACAGGG	480	
QY	481	CAGGGTGTCAGACAAGGATCGCGTGGGGCTATGATGATCTTGAGGCCCTTGATGGGTTTC	540	
Db	481	CAGGGTGTCAGACAAGGATCGCGTGGGGCTATGATGATCTTGAGGCCCTTGATGGGTTTC	540	
QY	541	TTCTTTGTATGTAACCTGGAAGGCAACTTGTGTGTTCACAGAAATGTGACACAGTAT	600	
Db	541	TTCTTTGTATGTAACCTGGAAGGCAACTTGTGTGTTCACAGAAATGTGACACAGTAT	600	
QY	601	CTAAGGTATTAACCAAGAAGAGCTGATGATGACAAAGGTATATAGCATTTGATGTTGGG	660	
Db	601	CTAAGGTATTAACCAAGAAGAGCTGATGATGACAAAGGTATATAGCATTTGATGTTGGG	660	
QY	661	GACCAACAGGAATTTGTCAAAAAACCCTGCTCCCAAGCTATATAGTAATGGGGGATCTTGG	720	
Db	661	GACCAACAGGAATTTGTCAAAAAACCCTGCTCCCAAGCTATATAGTAATGGGGGATCTTGG	720	
QY	721	TCGGGGAACCTCCGAGGGGGAACGCCATACCTTAATTTGTCGATGTCGTAAGAAACCT	780	
Db	721	TCGGGGAACCTCCGAGGGGGAACGCCATACCTTAATTTGTCGATGTCGTAAGAAACCT	780	
QY	781	TTACCTGATTCAGAAGAGAGGGTCACTGATTAACACAGAGAGCTCATGAGAAATGAGAACT	840	
Db	781	TTACCTGATTCAGAAGAGAGGGTCACTGATTAACACAGAGAGCTCATGAGAAATGAGAACT	840	
QY	841	ATGCAAGTCTTCGCTGTCTCTCAACCAAGTCCATCAAGAAGAGAGAGATTTTGGAG	900	
Db	841	ATGCAAGTCTTCGCTGTCTCTCAACCAAGTCCATCAAGAAGAGAGAGATTTTGGAG	900	
QY	901	TCCTGCTTATTTGGCTGGCAGAAAGTTCCTCATGAAGAAAGACCAAGTTCTTCCCTCA	960	
Db	901	TCCTGCTTATTTGGCTGGCAGAAAGTTCCTCATGAAGAAAGACCAAGTTCTTCCCTCA	960	
QY	961	TCAGAAAGTTTATCTACTGCGCAGAGATTCCTCAAGGCAAGATCACTGCTCTGGATTCACG	1020	
Db	961	TCAGAAAGTTTATCTACTGCGCAGAGATTCCTCAAGGCAAGATCACTGCTCTGGATTCACG	1020	
QY	1021	ACCATGAGAGCAGCATGAACACAGGCTTGAGAGACTGTTAAGAAAGTGTATTCAGAG	1080	

Db	1021	ACCAATGAGACGACATGAAACCAACGAGCTGGGAGACCTGTGTAAAGAGGTGATTCAGAG	1080
Qy	1081	TTCCATGCGGAGCATGAGAGAGAAATGTGTGTCCTATGCTAAAGAGCATCATCATGAAGTA	1140
Db	1081	TTCCATGCGGAGCATGAGAGAGAAATGTGTGTCCTATGCTAAAGAGCATCATCATGAAGTA	1140
Qy	1141	CTGAGACAAAGGATTGGCATTCAGTCAAATTCATCGTTTTTCTTGTGTGATGGACACTCTT	1200
Db	1141	CTGAGACAAAGGATTGGCATTCAGTCAAATTCATCGTTTTTCTTGTGTGATGGACACTCTT	1200
Qy	1201	GTTGCTGCAAAAGGAAAGACAAACATCCGTTCTTCACAGTACTAAATGAACCTCAACTT	1260
Db	1201	GTTGCTGCAAAAGGAAAGACAAACATCCGTTCTTCACAGTACTAAATGAACCTCAACTT	1260
Qy	1261	GTAATATCTTACATATGCTTTCACAGAGACAGAAATGTGTGTATGAATCCGGATCTG	1320
Db	1261	GTAATATCTTACATATGCTTTCACAGAGACAGAAATGTGTGTATGAATCCGGATCTG	1320
Qy	1321	ACTGGACAAACGATGGGGAAGCCACTGAATTCGAATTGCTTACAGCCCTGCCATCAG	1380
Db	1321	ACTGGACAAACGATGGGGAAGCCACTGAATTCGAATTGCTTACAGCCCTGCCATCAG	1380
Qy	1381	GCCCTGTGCASTGGGAACCCAGGTCAAGACATGACCTTAGTAGCAATATATATTTTCCC	1440
Db	1381	GCCCTGTGCASTGGGAACCCAGGTCAAGACATGACCTTAGTAGCAATATATATTTTCCC	1440
Qy	1441	ATTAATGGCCCCAAGGAAACAAATGGGCAATGGCCATGGGAGGTTTGGTGTGGGGGA	1500
Db	1441	ATTAATGGCCCCAAGGAAACAAATGGGCAATGGCCATGGGAGGTTTGGTGTGGGGGA	1500
Qy	1501	ATGAACCATGTGTGAGCATGCAAGCAACCACTCCTCAGGGTAGTAACTATGACACTCAA	1560
Db	1501	ATGAACCATGTGTGAGCATGCAAGCAACCACTCCTCAGGGTAGTAACTATGACACTCAA	1560
Qy	1561	ATGAACAGCCCTTCACAAAGCAGCCCTGGCATGTAAATCCAGGACACCCACCTCCATGCTT	1620
Db	1561	ATGAACAGCCCTTCACAAAGCAGCCCTGGCATGTAAATCCAGGACACCCACCTCCATGCTT	1620
Qy	1621	TCACCAAGGCATGCGCATGAGACCCCTGGAGTGGGCTGGCAGCCCTCGAATCCACCCAGTCAG	1680
Db	1621	TCACCAAGGCATGCGCATGAGACCCCTGGAGTGGGCTGGCAGCCCTCGAATCCACCCAGTCAG	1680
Qy	1681	TTTTTCCCCTGCGAGGAAGCTTGCAATTCCTCCCTGTGGGAGTTTGGAGCAGCACAAGAAATAGC	1740
Db	1681	TTTTTCCCCTGCGAGGAAGCTTGCAATTCCTCCCTGTGGGAGTTTGGAGCAGCACAAGAAATAGC	1740
Qy	1741	CATAGTATATACCAACAGCTCCCTCATAATGCATTCAGGCCCTTAGCGAGGGGACAGGGGTC	1800
Db	1741	CATAGTATATACCAACAGCTCCCTCATAATGCATTCAGGCCCTTAGCGAGGGGACAGGGGTC	1800
Qy	1801	TCATATAGGTCATGTGGTGGCTTCACCAAGACCTTAAATAATGGGCAATTTGCATAAACTCCCCA	1860
Db	1801	TCATATAGGTCATGTGGTGGCTTCACCAAGACCTTAAATAATGGGCAATTTGCATAAACTCCCCA	1860
Qy	1861	GTTATATATGAATTCCTCCCCCACTCAGCAAGATGGGAAGCTTGAGCTCAAAAAGACTGTTT	1920
Db	1861	GTTATATATGAATTCCTCCCCCACTCAGCAAGATGGGAAGCTTGAGCTCAAAAAGACTGTTT	1920
Qy	1921	GGACTATATGGGGAGCCCTTGGAAGTACAACTGGCAAGCAGAGACAGCTGCCATCTT	1980
Db	1921	GGACTATATGGGGAGCCCTTGGAAGTACAACTGGCAAGCAGAGACAGCTGCCATCTT	1980
Qy	1981	GGAGAGCAAAAGSAAACAAATGACCCCAACCTGCCCCGGCGGTAGAGAGTGAAGAGACT	2040
Db	1981	GGAGAGCAAAAGSAAACAAATGACCCCAACCTGCCCCGGCGGTAGAGAGTGAAGAGACT	2040
Qy	2041	GACGGCGAGACACATGCTATGACAGCAAAAGGACACCAAACTCCTCAGCTGTGTAC	2100
Db	2041	GACGGCGAGACACATGCTATGACAGCAAAAGGACACCAAACTCCTCAGCTGTGTAC	2100
Qy	2101	ACCAAAATCTGATCAGATGGAGCCCTGCCCCCTTAGCCAGCTCTTGTGGATACAAACAA	2160

Db 2101 ACCAAATCTGATCAGATGAGCCCTGCGCCTTACCCAGCTCTTTGTGCGATACAAACAA 2160  
QY 2161 GACTCCACAGTAGCTTCCCTGGTTCGTGGGTCTACACATGGAACCTGCTCAAGAGAG 2220  
Db 2161 GACTCCACAGTAGCTTCCCTGGTTCGTGGGTCTACACATGGAACCTGCTCAAGAGAG 2220  
QY 2221 CATAAAAATTTTGCACAGACTCTTTCAGAGACAGAGTTCCCTGTGGACTTTGGCCAACTTA 2280  
Db 2221 CATAAAAATTTTGCACAGACTCTTTCAGAGACAGAGTTCCCTGTGGACTTTGGCCAACTTA 2280  
QY 2281 ACAGCACAAGCCACAGCAGCAAGACCTTAGCCAGAGATCCAGACACAGCTCTGTATCA 2340  
Db 2281 ACAGCACAAGCCACAGCAGCAAGACCTTAGCCAGAGATCCAGACACAGCTCTGTATCA 2340  
QY 2341 GAAGTGCATTTTAAACAAGACCGGTGAGCCCAAGAAGAAAGAGATGCACTCTTCCG 2400  
Db 2341 GAAGTGCATTTTAAACAAGACCGGTGAGCCCAAGAAGAAAGAGATGCACTCTTCCG 2400  
QY 2401 TATTTGCTAGATTAAGATGATTAATAAGATATTGGTTTACCGAGAAATACCCCAACTT 2460  
Db 2401 TATTTGCTAGATTAAGATGATTAATAAGATATTGGTTTACCGAGAAATACCCCAACTT 2460  
QY 2461 GAGGAGATGAGACAGTAAAGACAGATCCCTGCCAGTAAACAAAATTAATAGCAATGAAACT 2520  
Db 2461 GAGGAGATGAGACAGTAAAGACAGATCCCTGCCAGTAAACAAAATTAATAGCAATGAAACT 2520  
QY 2521 GAGAGAGAGAGATGAGCTTTGAGCTGTGAGCAGCCTGGCAGTGAAGTGAAGTGAAGTGA 2580  
Db 2521 GAGAGAGAGAGATGAGCTTTGAGCTGTGAGCAGCCTGGCAGTGAAGTGAAGTGAAGTGA 2580  
QY 2581 GAGAGATTTTGGATGATGATTTGACAGATAGTCAATTACACAGCTTTTCCAGACAGAG 2640  
Db 2581 GAGAGATTTTGGATGATGATTTGACAGATAGTCAATTACACAGCTTTTCCAGACAGAG 2640  
QY 2641 CCAGGCGCCCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2700  
Db 2641 CCAGGCGCCCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2700  
QY 2701 ACAGCTGAAAAACAGCCCTGTCACACCTGTTGAGCCAGAAAAACAGACTGCGAATTTC 2760  
Db 2701 ACAGCTGAAAAACAGCCCTGTCACACCTGTTGAGCCAGAAAAACAGACTGCGAATTTC 2760  
QY 2761 CAGAGCACTTTTAATAACCCACAGCAGGCACTGGGCACTGGGCACTGGGCACTGGGCACT 2820  
Db 2761 CAGAGCACTTTTAATAACCCACAGCAGGCACTGGGCACTGGGCACTGGGCACTGGGCACT 2820  
QY 2821 TTACCACTTTACATCAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2880  
Db 2821 TTACCACTTTACATCAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2880  
QY 2881 AACAGTAGTCCCTACTACAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 2940  
Db 2881 AACAGTAGTCCCTACTACAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 2940  
QY 2941 GGAACCAAGGAAATTTAGGGAACAGTACAGAGGAAATGATGATGATGATGATGATGAT 3000  
Db 2941 GGAACCAAGGAAATTTAGGGAACAGTACAGAGGAAATGATGATGATGATGATGATGAT 3000  
QY 3001 CCACTATGCTATGCTGAGAGATGGGCAAGCAGGCTGCTGCTGCTGCTGCTGCTGCTGCT 3060  
Db 3001 CCACTATGCTATGCTGAGAGATGGGCAAGCAGGCTGCTGCTGCTGCTGCTGCTGCTGCT 3060  
QY 3061 GCTACCAACAGTGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3120  
Db 3061 GCTACCAACAGTGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3120  
QY 3121 AGCATCCCATGAGGCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 3180  
Db 3121 AGCATCCCATGAGGCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 3180  
QY 3181 ATGAATATAGGGCATCTGAAATTAAGATGATGATGATGATGATGATGATGATGATGAT 3240  
Db 3181 ATGAATATAGGGCATCTGAAATTAAGATGATGATGATGATGATGATGATGATGATGAT 3240

QY 3241 CAACTCCCTCCAAATCAGACTGCCCCCATGGGCTGAAGACATCCCTGCTATAGACAGAGG 3300  
Db 3241 CAACTCCCTCCAAATCAGACTGCCCCCATGGGCTGAAGACATCCCTGCTATAGACAGAGG 3300  
QY 3301 TCTTTTCCAGCCAAAACAGGACAGCCATTTTGGCAGTCTCCAGATGATGATGATGATGAT 3360  
Db 3301 TCTTTTCCAGCCAAAACAGGACAGCCATTTTGGCAGTCTCCAGATGATGATGATGATGAT 3360  
QY 3361 CATCTCCAGCTGAGTCTCCAGATGATGATGATGATGATGATGATGATGATGATGATGAT 3420  
Db 3361 CATCTCCAGCTGAGTCTCCAGATGATGATGATGATGATGATGATGATGATGATGATGAT 3420  
QY 3421 TTGCGGAATTTTGAATGAGCTGGAGAGATGATGATGATGATGATGATGATGATGATGAT 3480  
Db 3421 TTGCGGAATTTTGAATGAGCTGGAGAGATGATGATGATGATGATGATGATGATGATGAT 3480  
QY 3481 AGCCAGAGCCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3540  
Db 3481 AGCCAGAGCCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3540  
QY 3541 GAGCAGAAGGCGCCGCTTTTCCACAGCAGTATGATGATGATGATGATGATGATGATGAT 3600  
Db 3541 GAGCAGAAGGCGCCGCTTTTCCACAGCAGTATGATGATGATGATGATGATGATGATGAT 3600  
QY 3601 AGCTATTCCTCCATGACAGATGATGATGATGATGATGATGATGATGATGATGATGAT 3660  
Db 3601 AGCTATTCCTCCATGACAGATGATGATGATGATGATGATGATGATGATGATGATGAT 3660  
QY 3661 ACACCTCGTATGACAGCCAGACCGGCTCAGGCCCCAGGCCCCAGGCCCCAGGCCCCAG 3720  
Db 3661 ACACCTCGTATGACAGCCAGACCGGCTCAGGCCCCAGGCCCCAGGCCCCAGGCCCCAG 3720  
QY 3721 AATCACTAAGACTTCAACTTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 3780  
Db 3721 AATCACTAAGACTTCAACTTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 3780  
QY 3781 ATGAATCAAAATCAGCAATGTTTCCAAATGATGATGATGATGATGATGATGATGATGAT 3840  
Db 3781 ATGAATCAAAATCAGCAATGTTTCCAAATGATGATGATGATGATGATGATGATGATGAT 3840  
QY 3841 CAGGACCTATTAATGACAGATGCTGCGCCAGAGACAGAGGGAATCTGAACAGCAT 3900  
Db 3841 CAGGACCTATTAATGACAGATGCTGCGCCAGAGACAGAGGGAATCTGAACAGCAT 3900  
QY 3901 CTTGAGACAGACAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3960  
Db 3901 CTTGAGACAGACAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3960  
QY 3961 GGACAAAGGTTGAATGACAGCAAGATGATGATGATGATGATGATGATGATGATGATGAT 4020  
Db 3961 GGACAAAGGTTGAATGACAGCAAGATGATGATGATGATGATGATGATGATGATGATGAT 4020  
QY 4021 AGCAACCCCTGGATTTCCCGAGGCAATGACAGAGTTTCCATTTCTCCAACTACGA 4080  
Db 4021 AGCAACCCCTGGATTTCCCGAGGCAATGACAGAGTTTCCATTTCTCCAACTACGA 4080  
QY 4081 ATAAATGACAGCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4140  
Db 4081 ATAAATGACAGCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4140  
QY 4141 TCACCCCGAATGACATGACAGAGATGATGATGATGATGATGATGATGATGATGATGAT 4200  
Db 4141 TCACCCCGAATGACATGACAGAGATGATGATGATGATGATGATGATGATGATGATGAT 4200  
QY 4201 TATCAGGCCCCCTCCGACATTAATGATGATGATGATGATGATGATGATGATGATGAT 4260  
Db 4201 TATCAGGCCCCCTCCGACATTAATGATGATGATGATGATGATGATGATGATGATGAT 4260  
QY 4261 TTTTCCAGAGATGCTCCACACACTTTGGGCAAGCAACACAGCATGATGATGATGAT 4320  
Db 4261 TTTTCCAGAGATGCTCCACACACTTTGGGCAAGCAACACAGCATGATGATGATGATGAT 4320

QY 4321 AACATGACATCAATGTGTCTCATGTGGCCAAACACAGTGGCATGAGCAGATGAACAG 4380  
|||||  
Db 4321 AACATGACATCAATGTGTCTCATGTGGCCAAACACAGTGGCATGAGCAGATGAACAG 4380  
QY 4381 ATGACAGGACAGATGACGATGACCTGACCTCGGTGTCTACGTACAGGCTGTCTCC 4440  
|||||  
Db 4381 ATGACAGGACAGATGACGATGACCTGACCTCGGTGTCTACGTACAGGCTGTCTCC 4440  
QY 4441 ATGGGTCCGAGCAGGTATATGATCTGTGCTGAGGGAGGAGCAACTGTTCCCAACAG 4500  
|||||  
Db 4441 ATGGGTCCGAGCAGGTATATGATCTGTGCTGAGGGAGGAGCAACTGTTCCCAACAG 4500  
QY 4501 CTGGCTGAGATGAGATGATTAAGCAGGAGGAGACACAACAGGAATATTGCTGACAC 4560  
|||||  
Db 4501 CTGGCTGAGATGAGATGATTAAGCAGGAGGAGACACAACAGGAATATTGCTGACAC 4560  
QY 4561 TGTGTGAAGCAGTGTGTCTTCTGACGTGACCGGGCTACTTGTCTCAAAACACTTCCAGTCT 4620  
|||||  
Db 4561 TGTGTGAAGCAGTGTGTCTTCTGACGTGACCGGGCTACTTGTCTCAAAACACTTCCAGTCT 4620  
QY 4621 GGAGAGCTGTGTCTATTTTTCACCCCAACTGACCTGGCAGCGGGTCTGTGAGACAG 4680  
|||||  
Db 4621 GGAGAGCTGTGTCTATTTTGTTCACACCACACTGACCTGGCAGCGGGTCTGTGAGACAG 4680  
QY 4681 ACAGGCTGGCCCTGGTTCACAGGGTGGCGTCCACTCGGCTGTGGCAGGAGAGCTGCT 4740  
|||||  
Db 4681 ACAGGCTGGCCCTGGTTCACAGGGTGGCGTCCACTCGGCTGTGGCAGGAGAGCTGCT 4740  
QY 4741 CTCTCTTACAGCTGAGAGCTTCGACATCCAGACAGTCCGCTGCTGCTCCGACTTCA 4800  
|||||  
Db 4741 CTCTCTTACAGCTGAGAGCTTCGACATCCAGACAGTCCGCTGCTGCTCCGACTTCA 4800  
QY 4801 CCTTATGCACTTATAGATCTCTCCCAAGTAATGTTGACAGGCCATTTCAATACC 4860  
|||||  
Db 4801 CCTTATGCACTTATAGATCTCTCTCCCAAGTAATGTTGACAGGCCATTTCAATACC 4860  
QY 4861 ATGTCAAGTGAATGTATTTAAATGATGATTAAGGAGAACCAATGCTTGTCTGT 4920  
|||||  
Db 4861 ATGTCAAGTGAATGTATTTAAATGATGATTAAGGAGAACCAATGCTTGTCTGT 4920  
QY 4921 CCTGTGCTGCTCAGACACCTGGTCTGTGCTTGTCCGCTGCTACAGCTACTGCTC 4980  
|||||  
Db 4921 CCTGTGCTGCTCAGACACCTGGTCTGTGCTTGTCCGCTGCTACAGCTACTGCTC 4980  
QY 4981 AAAGATTAAGATTTATCTGCGGGGAAAGAAAGATTTTAAATAATTAAGTAAGTAAG 5040  
|||||  
Db 4981 AAAGATTAAGATTTATCTGCGGGGAAAGAAAGATTTTAAATAATTAAGTAAGTAAG 5040  
QY 5041 GTTTTAAGCTAAAGCCTGATTTGGGATGGAAGCAGACAGACCGGTGACAGCGCTGT 5100  
|||||  
Db 5041 GTTTTAAGCTAAAGCCTGATTTGGGATGGAAGCAGACAGACCGGTGACAGCGCTGT 5100  
QY 5101 ATTACAGACACACCCAGCGCTGAAGACCAACAAGTCACTGTATCTCTAGAAAGC 5160  
|||||  
Db 5101 ATTACAGACACACCCAGCGCTGAAGACCAACAAGTCACTGTATCTCTAGAAAGC 5160  
QY 5161 TCTAAAGACCATGTTGGAAGAAGTCTCAGTACTGAACAGATGGAAGAAGAGCCTGTGAG 5220  
|||||  
Db 5161 TCTAAAGACCATGTTGGAAGAAGTCTCAGTACTGAACAGATGGAAGAAGAGCCTGTGAG 5220  
QY 5221 AGGGCTGTTAACATTTGCAAAATTTTTCCTGTTTCTTTTCTTTTAAACCAACTGG 5280  
|||||  
Db 5221 AGGGCTGTTAACATTTGCAAAATTTTTCCTGTTTCTTTTCTTTTAAACCAACTGG 5280  
QY 5281 TTCACTGATCATGTAATGAGAGAAATTAATTCTATTCTAATTAAGTCCCTTTTGA 5340  
|||||  
Db 5281 TTCACTGATCATGTAATGAGAGAAATTAATTCTATTCTAATTAAGTCCCTTTTGA 5340  
QY 5341 TTTGATCAGACAGCTTGAATTCACATCTCTTCTCCCTGACAGCCTGACTCTTCCCTTC 5400  
|||||  
Db 5341 TTTGATCAGACAGCTTGAATTCACATCTCTTCTCCCTGACAGCCTGACTCTTCCCTTC 5400  
QY 5401 CCTCTCTCATTTCCCACTCTCTATTTTCATTTCTTTTAAATAATTAAGCTAC 5460  
|||||

|||||  
Db 5401 CCTCTCTCATTTCCCACTCTCTATTTTCATTTCTTTTAAATAATTAAGCTAC 5460  
QY 5461 AGAACCAGGTAGGCCCTTATTTCCCTTAATGTTTGGCAGCCATTAACCAATGTCTAA 5520  
|||||  
Db 5461 AGAACCAGGTAGGCCCTTATTTCCCTTAATGTTTGGCAGCCATTAACCAATGTCTAA 5520  
QY 5521 GTATTGAATTTGAGAAAAAATGATTTACTGCAAGGAGAGCAAGTAAGTAAAGC 5580  
|||||  
Db 5521 GTATTGAATTTGAGAAAAAATGATTTACTGCAAGGAGAGCAAGTAAGTAAAGC 5580  
QY 5581 TTGATACCAATGAGCTTAAGATTAACCTGCTTGAAGCATGTTATTTCTGTTCCACGA 5640  
|||||  
Db 5581 TTGATACCAATGAGCTTAAGATTAACCTGCTTGAAGCATGTTATTTCTGTTCCACGA 5640  
QY 5641 ACTCTGGCTCCAAATGGAGAAACGCCAGTGTGTTTAAATTTGATPAGCATATCAGA 5700  
|||||  
Db 5641 ACTCTGGCTCCAAATGGAGAAACGCCAGTGTGTTTAAATTTGATPAGCATATCAGA 5700  
QY 5701 CAGATTTAACTCTGCAATGTGTTTATTTTGTGTTTAACTGCTGACTAAGCCGA 5760  
|||||  
Db 5701 CAGATTTAACTCTGCAATGTGTTTATTTTGTGTTTAACTGCTGACTAAGCCGA 5760  
QY 5761 AGTTTGTAAAGTACATAAAATCAATTTATGTAACACCAATTAATTTAAGTTGAGA 5820  
|||||  
Db 5761 AGTTTGTAAAGTACATAAAATCAATTTATGTAACACCAATTAATTTAAGTTGAGA 5820  
QY 5821 ACTTATGCTTTTAAATGATTAATTTTGTGAGTATACATTTGGAATTTGACTCAAA 5880  
|||||  
Db 5821 ACTTATGCTTTTAAATGATTAATTTTGTGAGTATACATTTGGAATTTGACTCAAA 5880  
QY 5881 AATGAGTACTTCAGTATTAATTAATGATATCTTCAATAGCAATGTCTTAAAGTGTCT 5940  
|||||  
Db 5881 AATGAGTACTTCAGTATTAATTAATGATATCTTCAATAGCAATGTCTTAAAGTGTCT 5940  
QY 5941 GTAAAGATATCAATGCTGTTGATTAAGCTAATTTGTACACTTAAGCTTTTATTTCT 6000  
|||||  
Db 5941 GTAAAGATATCAATGCTGTTGATTAAGCTAATTTGTACACTTAAGCTTTTATTTCT 6000  
QY 6001 AAACCTTGATTTGCTATATAGTATATCTATATCTATATGATATGATGACGCTGTA 6060  
|||||  
Db 6001 AAACCTTGATTTGCTATATAGTATATCTATATCTATATGATATGATGACGCTGTA 6060  
QY 6061 GGAACCAATTCGATTTTATATGTTTATATCTTCTTAAATGAACCTTAGAAAGACTA 6120  
|||||  
Db 6061 GGAACCAATTCGATTTTATATGTTTATATCTTCTTAAATGAACCTTAGAAAGACTA 6120  
QY 6121 CATGTTACTAAGCAGGCCACTTTTATGTTGTTTTT 6156  
|||||  
Db 6121 CATGTTACTAAGCAGGCCACTTTTATGTTGTTTTT 6156

## RESULT 2

US-09-723-535-3  
; Sequence 3, Application US/09723535  
; Patent No. 635483  
; GENERAL INFORMATION:  
; APPLICANT: C. Frank Bennett  
; APPLICANT: Lex M. Cowert  
; TITLE OF INVENTION: ANTISENSE MODULATION OF SRC-2 EXPRESSION  
; FILE REFERENCE: RUS-0225  
; CURRENT APPLICATION NUMBER: US/09/723,535  
; NUMBER OF SEQ ID NOS: 49  
; SEQ ID NO. 3  
; LENGTH: 6156  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (163)..(4557)  
US-09-723-535-3



||||| 2160  
Db accaaatcgcatacagatgagccctcgcccttagccagctcttctgcgatacaacaaca 2160  
QY 2161 GACTCCACAGTAGCTTGGCTGGTTCTGGTCTACACATGGAACCTCGCTCAAGAGAG 2220  
2161 gactccacagtagcttgcctcgcttctgcgatacaacaacaacaacaagaag 2220  
QY 2221 CATAAATTTTGCACAGACTCTTGCAGACAGCAGTTCCTGTGACTTGGCCAATT 2280  
2221 cataaatttgcacagactcttgcagcagcagagcttccctctgtgacttgcgaagta 2280  
Db 2281 ACAGCAAGAACCCACAGGCAAAAGACCTGACCGAGATCCAGACGACACTCTGTGATCA 2340  
2281 acagcagaaagccacagcagaagaccctgagccagagctccagcagcagctctgata 2340  
QY 2341 GAAGTACTATTAAACAAAGCGCGTGAGCCCCCAAGAAAGAGATGACTACTTCCG 2400  
2341 gaagtactattaaacaagaagccggtgagcccaagaagaagaatgcaactctgcg 2400  
Db 2401 TATTGCTGATTAAGATGATCTAATAAGATATTGTTTACCAAAATAACCCCAACTT 2460  
2401 tatttctgataaagatgataactaaagatattgttaccagaataaacccccaaact 2460  
QY 2461 GAGAGACTGACAGTAAAGACAGATCCTGCCAGTAAACAAAATTATAGCAATGAAACT 2520  
2461 gagagactgacagtaagacagatcctgcagtaacacaaatlaaagaatgaaact 2520  
Db 2521 GAGAAAGAGAGATGAGCTTTGAGCTGTGAGACAGCTGGCAGTGAAGTGGCAACTTG 2580  
2521 gagaagaagagatgagcttgcagctgtgagccagccgagcagtgagctggcaacttg 2580  
QY 2581 GAGGAGATTTTGGATGATTTGCAAGATAGTCAATTACACAGCTTTTCCAGACAGAG 2640  
2581 gaggagatttggatgatttgcagaatagtaacaaattaccagacttcccaagacagag 2640  
Db 2641 CCAGGCGCCCTGCTGATCAGTTGACAGCAAGCCATATCATATACCTCATGCAACTC 2700  
2641 ccaggcgccctgctgatactgatacgaagcaagcatacaatgacccatgcaactc 2700  
QY 2701 ACAGCTGAAAACAGCCCTGTCAACACTGTGAGCCAGAAAACACACTGCGAATTTCA 2760  
2701 acagctgaaaaacagccctgtcaacctgttggagccagaaaaacagactcggaatttca 2760  
Db 2761 CAGAGCACTTTTAAATACCCACAGCAGGGAACCTGGGCAAGTTATGCCAAACAGAA 2820  
2761 cagagcacttttaataaccacagcagcagcagcagctgagcttattctgccaacccagaa 2820  
QY 2821 TTACCACTTGACATCATTTGCAAGCCCAACTGGTGGAGACTTTCACCAATCAGA 2880  
2821 ttaccacttgacatcatttgcaaaagcccaactggtctgagaccttcccaaccaatcaga 2880  
Db 2881 AACAGTAGTCCCTACTCAGTATGATCTCAGCCAGCAAGATGAGGTATCAAGGATGAT 2940  
2881 aaagtagtccctactcactcagtatactctcagcaggaatgataatgatacaaggaatgata 2940  
QY 2941 GGAACCAAGAAATTTAGGGAACAGTAGCAGCAAGATGTTGTAACAGTCTTCG 3000  
2941 ggaaccaagaaatttaggaacagtagcacagaatgattgtaaacagtgcttctcgg 3000  
Db 3001 CCTACTGCACTTGAGAGATGGGCAACCGCAGATTCGGCTGTGAGAGTCACTGTGCT 3060  
3001 cctactgcacttgagagaatgagcagcagagatctggctgtgagagtcacacttgct 3060  
QY 3061 GGTACACACAGTGCATGAAACCGGCAGTCCAGAGAGGTGATTGGAAACCCAGAGCC 3120  
3061 ggtacacacagtgcattgaacccgagctccaaaggaagtatgattggaacccaagcagcc 3120  
Db 3121 AGCATCCCATGAGGCCACAGCAGCAGCAGCTGGCCAAAGACAGAGCTTCACTGCTCAG 3180  
3121 agcatcccatgagggccacagcagcagcagcagcagcagcagcagcagcagcagcagcag 3180  
QY 3181 ATGAATATATAGGCGCATCTGATTAAGAGATGACATGGGGGACCTCATGTTACGCAACA 3240  
3181 atgaatatatagggccatctgaaatagagatgaaatgaggggagcctcagtatagccaaca 3240  
Db 3241 CAAGCTCTCTCAAAATCAGACTGCCCATGGCTGAAAGCATCTGCTATAGACCAAGGCG 3300  
3241 caagctctctcaaaatcagacttgcctccatgctgaaagaatctctcctatagaccagagc 3300  
QY 3301 TCTTTTGGCAGCCAAAACAGGCAAGCCATTGGCAGTTCTCCAGATGACTTGTATGTCCA 3360  
3301 tcttttggcagccaaaacagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 3360  
Db 3361 CATCCGTCAGTCTGAGTCTCGATGATGAGGAGGCTCTCTGACAGCTGATCTATCGGCC 3420  
3361 catccgtcagctgagcttccgagtgatgagagagctctcttgaaacagctgatactgccc 3420  
QY 3421 TTGCGAATTTTGAATGGCTTGAGAGAGATGATGAGCTTGAATATCCCAACTGTGTC 3480  
3421 ttgcgaaatttgaatggcttgagagagatgattgagagcttgaatataccgaactgttc 3480  
Db 3481 AGCCAGAGCCAAAGCAGTATGATCCAGAACAGTTCTCAAGTCAGAGATTCCAACTCATGCTG 3540  
3481 agccagagccaaagcagtagatccagacagttctcaagtcagagatccacaatcatgctg 3540  
QY 3541 GAGCAAGAGCGCGCTTTTCCACAGCAGTATGATCTCAGGCAACAAATGGCCAGGGT 3600  
3541 gagcaagagcgcgcttttccacagcagatgatactcacaacatggagcagggctgatactgc 3600  
Db 3601 AGCTATCTCCCATGCAAGATCCAAACTTTCACACCAGTGGGAGCAGCGGCTATGATGCC 3660  
3601 agctatctcccatgcaagatccaaacttccacaacatggagcagggctgatactgc 3660  
QY 3661 ACACCTCGTATGACAGCCGACAGCGGCTTCAGGCCCAGCGGCTTATGAGAAACAGGCA 3720  
3661 acacctcgtatgacagccgacagcggcttcagggcccaagcagggcttgaagaacagcaga 3720  
Db 3661 acactcgtatgacagcgagcagcggcttcagggcccaagcagggcttgaagaacagcaga 3720  
QY 3721 AATCACTTAAGACTTCACTTACCTTGAGCTTCAGCTTCACAGCAGAGCAATGCGCAGCCACTT 3780  
3721 aatcaacttaagacttcaacttgaacttgaacttgaacttgaacttgaacttgaacttgaacttga 3780  
Db 3721 aatcaacttaagacttcaacttgaacttgaacttgaacttgaacttgaacttgaacttgaacttga 3780  
QY 3781 ATGAATCAATCAGCAATGTTTCCAAATGTAAGTGAATGAACTTGAAGCTGTGAGGCTGAGTACCAACA 3840  
3781 atgaatcaatcagcaatgtttccaaatgtaagtgaatgaaacttgaagctgtgaggtacccaaca 3840  
Db 3781 atgaatcaatcagcaatgtttccaaatgtaagtgaatgaaacttgaagctgtgaggtacccaaca 3840  
QY 3841 CAGGCACTTATTAATGACAGATGCTGGGCCAGAGACAGAGAAATCTGAAACAGCAT 3900  
3841 caggcacttattatgacagatgctgggccagagacagagaaatctgaaacagcat 3900  
Db 3841 caggcaacttattatgacagatgctgggccagagacagagaaatctgaaacagcat 3900  
QY 3901 CTTCGACAGAGACAAATGATCAGCAAGCAAGCAATGTTTCAAGCAAGCAATGATGATGAGA 3960  
3901 cttcgacagagacaaatgatcagcaagcaagcaagcttcagcaagcaagcaagcttgaatgagaga 3960  
Db 3901 cttcgacagagacaaatgatcagcaagcaagcaagcttcagcaagcaagcaagcttgaatgagaga 3960  
QY 3961 GGAACAAGGTTGAATATGACACCAAGCATGAGGCTCTAGTGTATGACCAACTATG 4020  
3961 ggaacaaggttgaatatgacaccaagcatgaggtctctagtgatgacccaactatg 4020  
Db 3961 ggaacaaggttgaatatgacaccaagcatgaggtctctagtgatgacccaactatg 4020  
QY 4021 AGCAACCTTCGGATTTCCCAAGGCAATGCAAGCAATGTTTCAATTTCTTCCAACTACGGA 4080  
4021 agcaaccttcggatttcccaaggcaatgcaagcaatgtttcaatttcttccaaactacgga 4080  
Db 4021 agcaaccttcggatltcccaggaatgcaagcaagtttccatlttccccaactacgga 4080  
QY 4081 ATAAATCAGCAACCTGATCAGAGCTTACTGAGGCTTCAGCACTCCAGAGCCCACTTATG 4140  
4081 ataaatcagcaacctgatcagagcttactgaggttcagcaactcccaagcccaactatg 4140  
Db 4081 ataaatcagcaacctgatcagagcttactgaggttcagcaactcccaagcccaactatg 4140  
QY 4141 TCACCCCGAATGGCAGATACACAGAGTCCCATGATGATCAACAGTCTCAGGCAACCCAGCC 4200  
4141 tcaccccgaaatggcagatcacagagttccatgataigcaacagctcagccaacccaagcc 4200  
Db 4141 tcaccccgaaatggcagatcacagagttccatgataigcaacagctcagccaacccaagcc 4200  
QY 4201 TATCAGGCCCCCTCCGACATTAATGATGGGCGCAGGGAAATGCGGCGGAAACACACTAG 4260  
4201 tatcagggccccctccgacataatgataatgagcaggggaataatgagcaggaacagcagcag 4260  
Db 4201 tatcagggccccctccgacataatgataatgagcaggggaataatgagcaggaacagcagcag 4260  
QY 4261 TTTTCCAGAGAGTCCGACACACTTGGGCAAGCAAGCAAGCAAGCAAGTACAGTATAC 4320  
4261 ttttccagagagttccgacacacttgggcaagcaagcaagcaagcaagcagcagcagcagcagcag 4320  
Db 4261 ttttccagagagttccgacacacttgggcaagcaagcaagcaagcagcagcagcagcagcagcagcag 4320



QY 4321 AACATGAACATCAATGTGTCCATGCGACCAACAGGTGGCATGAGCATGAACACG 4380  
|||||  
Db 4321 aacatgaacatcaatgtgtccatgcgaccaaacaggtggcatgagcatgaacacg 4380  
QY 4381 ATGACGAGACAGATCAGATGACCTGACGACCTGCTGCTACGAGGCTGTCTCC 4440  
|||||  
Db 4381 atgacgagacagatcagatgacacctgacgacctgctgctacgagggctgtctcc 4440  
QY 4441 ATGGGTCCCGAGCAGATTAATGATTCCTGCTGAGGGGAGGCAACCTGTCCCAACG 4500  
|||||  
Db 4441 atgggtcccgagcagatgataatgctcctgaggggagggcaaacctgtcccaacg 4500  
QY 4501 CTGCGTGAATGATGATGATTAAGCAGAGGAGACACACGGAATTAATGATGACAC 4560  
|||||  
Db 4501 ctgctggaatgatatgatatgataagagagagagacacacggaaatataatgacac 4560  
QY 4561 TGCTGAAGCCAGTGTGCTTCTCAGCTGACCGGGCTCACTTGCTCAAAACATTCAGCT 4620  
|||||  
Db 4561 tgcctgaagccagtgtgctctcttcacgtgacgggtctcactgtcctcaaacatctcagct 4620  
QY 4621 GGAGAGCTGTGTCTATTTGTTTCAACCACTGACCTGCGACCGGTTCTGCTAGACAG 4680  
|||||  
Db 4621 ggagagctgtgtctatctgtctcaaaccaactgacctgacgggtctcgttagagcag 4680  
QY 4681 ACAGGCTGGGCTGGTTCACAGGCTGGGCTGACCTGGGCTGGGCTGGGCTGGGCT 4740  
|||||  
Db 4681 acagggctgggctgggttcacagggctgggctgacacctgggctgggctgggctgggct 4740  
QY 4741 CTCTCTGTGACAGTCTGAGCTGCGACCTGACAGCTGCTGCTGCTGCTGCTGCTGCT 4800  
|||||  
Db 4741 ctctctgtgacagctgacagctgacagctgacagctgacagctgacagctgacagct 4800  
QY 4801 CCTTATGTCACATTAATGATCTCTCTCCCAAGTAATGATGACAGGCAATTTCAATCC 4860  
|||||  
Db 4801 ccttatgtcacatgataatgataatgataatgataatgataatgataatgataatgata 4860  
QY 4861 ATGTCAGATGATGATGATTAATGATGATTAATGATGATTAATGATGATTAATGAT 4920  
|||||  
Db 4861 atgtcagatgatatgatatgatatgatatgatatgatatgatatgatatgatatgatat 4920  
QY 4921 CCTGTGCGTTCAGACACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4980  
|||||  
Db 4921 cctgtgctgttcagacactgctgctgctgctgctgctgctgctgctgctgctgct 4980  
QY 4981 AAAATTAAGATTTTATGTTGGGGGAAAGAAAGAAATTTTAAATTAATTAATTA 5040  
|||||  
Db 4981 aaatattaagatttatactggtgggaaagaaagaaatltttaaaaaatlaaactaaagat 5040  
QY 5041 GTTTTAAAGCTTAAGCTTAATTTGGAGTGAAGACAGACACGCTGGACAGCTGT 5100  
|||||  
Db 5041 gttttaagcttaagcttaaatggtgagagagagagacacacgctggacagctgtgt 5100  
QY 5101 ATTTACAGACACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5160  
|||||  
Db 5101 atttacagacacacccctgctgctgctgctgctgctgctgctgctgctgctgctgct 5160  
QY 5161 TCTTAAAGACCTGTTGGAAAGAGTCTCAGTTACTGAACAGATGAAAGAGCTGTAG 5220  
|||||  
Db 5161 tctttaaagacctgttggaaaagagctcagttactgaacagatgaaaagagctgttag 5220  
QY 5221 AGGCGTGTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAA 5280  
|||||  
Db 5221 aggcgtgttaacttaacttaacttaacttaacttaacttaacttaacttaacttaact 5280  
QY 5281 TTACCTGATCATGATTTGAAGAAATTTTCAATTTCAATTTCAATTTCAATTTCA 5340  
|||||  
Db 5281 ttacctgatcatgatttgaagaaattttcaattttcaattttcaattttcaattttca 5340  
QY 5341 TTGATTCAGACAGCTTGAATGACATCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5400  
|||||  
Db 5341 ttgatctcagacagcttgaatgacatctctctctctctctctctctctctctctctcc 5400

QY 5401 CCTCTGATTTCCCATCTCCCTATTTTCAATCTCTCTCTCTCTCTCTCTCTCTCT 5460  
|||||  
Db 5401 cctctctgatttcccatctccctatttctctctctctctctctctctctctctctctcc 5460  
QY 5461 AGAAGCAGATTAAGCTTATTTCTTAAATGTTTGGCAGCCACTTACCAATTTGCT 5520  
|||||  
Db 5461 agaagcagatgagccttattctcttaaatglttggcagccacttaaccattgctaa 5520  
QY 5521 GATTTGAATTTGAGAAAAAATGATTTCTGCGAAGGAGGAGGAGGAGGAGGAGG 5580  
|||||  
Db 5521 gatattaatttgaagaaaaaatgatttctgcgaaaggagagagagagagagagagag 5580  
QY 5581 TTGATACCAATGACATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTA 5640  
|||||  
Db 5581 ttgataccaatgacattaaagattaaagattaaagattaaagattaaagattaaagatt 5640  
QY 5641 ACTTGCCCTCCAAAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 5700  
|||||  
Db 5641 acttgccctccaaaatgag 5700  
QY 5701 CAGATTTAACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5760  
|||||  
Db 5701 cagatttaacctctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 5760  
QY 5761 AGTTTGAAGTACATAAATCAATTTATATGTAACCAATTAATTTAACTTAACT 5820  
|||||  
Db 5761 agtttgaagtacataaatacaatttatatgtaaaccaatataatgtaagttgagag 5820  
QY 5821 ACTTATGTTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAA 5880  
|||||  
Db 5821 acttatgttttaattttaaattttaaattttaaattttaaattttaaattttaaattt 5880  
QY 5881 AATGAGTACTTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5940  
|||||  
Db 5881 aatgagtacttcaattataatgataatgataatgataatgataatgataatgataat 5940  
QY 5941 GTAAGATATCAATGCTTGAATTAATTAATTAATTAATTAATTAATTAATTA 6000  
|||||  
Db 5941 gtaagatatcaatgcttgatgataatgataatgataatgataatgataatgataat 6000  
QY 6001 AAACCTTGATTTGCTTATTAATTAATTAATTAATTAATTAATTAATTAATTA 6060  
|||||  
Db 6001 aaaccttgatctgtctataatgataatgataatgataatgataatgataatgataat 6060  
QY 6061 GGAACCAATTTCTGATTTTATTAATTAATTAATTAATTAATTAATTAATTA 6120  
|||||  
Db 6061 ggaaccaatctgtatttataatgataatgataatgataatgataatgataatgataat 6120  
QY 6121 CATGTTACTAAGCAGGCACTTTATGTTGTTT 6156  
|||||  
Db 6121 catgttactaagcagcactttatggtgtttt 6156

RESULT 3  
US-09-723-534-3  
Sequence 3, Application US/09723534  
Patent No. 6294382  
GENERAL INFORMATION:  
APPLICANT: C. Frank Bennett  
TITLE OF INVENTION: ANTISENSE MODULATION OF SRC-1 EXPRESSION  
FILE REFERENCE: RFS-0225  
CURRENT FILING DATE: 2000-11-27  
NUMBER OF SEQ ID NOS: 49  
SEQ ID NO 3  
LENGTH: 4547  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (43) .. (4368)  
US-09-723-534-3



Db 633 CTCGAGGTCCTAGAGCACT 653

```

5
US-08-785-310A-1
Sequence 1, Application US/08785310A
Patent No. 5840532
GENERAL INFORMATION:
APPLICANT: McKnight, Steven L.
APPLICANT: Russell, David W.
TITLE OF INVENTION: Neuronal PAS Domain Protein
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
STREET: 268 BUSH STREET, SUITE 3200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/785,310A
FILING DATE: 21-JAN-1997
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, RICHARD A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: UTSD:1226
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 343-4341
TELEFAX: (415) 343-4342
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2078 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-785-310A-1

```

[illegible]

RESULT 6  
US-08-885-291-54  
; Sequence 54, Application US/08885291A  
; Patent No. 6057125  
; GENERAL INFORMATION:  
; APPLICANT: Takahashi, Joseph S.

```

1  APPLICANT: Turek, Fred W.
2  APPLICANT: Pinto, Lawrence H.
3  TITLE OF INVENTION: CLOCK GENE AND GENE PRODUCT
4  FILE REFERENCE: 0250-5
5  CURRENT APPLICATION NUMBER: US/08/885,291A
6  CURRENT FILING DATE: 1997-06-30
7  EARLIER APPLICATION NUMBER: 08/816,693
8  EARLIER FILING DATE: 1997-03-13
9  NUMBER OF SEQ ID NOS: 55
10 SOFTWARE: PatentIn Ver. 2.0
11 SEQ ID NO 54
12     LENGTH: 3545
13     TYPE: DNA
14 ORGANISM: Homo sapiens
15 US-08-885-291-54

```

[illegible]

```

1  RESULT 7
2  US-09-496-672--54
3  Sequence 54, Application US/09496672
4  Patent No. 6291429
5  GENERAL INFORMATION:
6  APPLICANT: Takahashi, Joseph S.
7  APPLICANT: Turek, Fred W.
8  APPLICANT: Plinto, Lawrence H.
9  TITLE OF INVENTION: CLOCK GENE AND GENE PRODUCT
10 FILE REFERENCE: 0290-5
11 CURRENT APPLICATION NUMBER: US/09/496,672
12 CURRENT FILING DATE: 2000-02-03
13 PRIOR APPLICATION NUMBER: 08/885,291
14 PRIOR FILING DATE: 1997-06-30
15 PRIOR APPLICATION NUMBER: 08/916,693
16 PRIOR FILING DATE: 1997-03-13
17 NUMBER OF SEQ ID NOS: 55
18 SOFTWARE: PatentIn Ver. 2.0
19 SEQ ID NO 54
20 LENGTH: 3545
21 TYPE: DNA
22 ORGANISM: Homo sapiens
23 US-09-496-672--54

```

	Query Match	0.9%	Score 55	DB 4	Length 3545
	Best Local Similarity	55.5%	Pred. NO. 2.8e-05		
	Matches 106	Conservative	0	Mismatches 85	Indels 0
					Gaps 0
QY	496	AAGATGCGCGTGGGCGCTATGATGCTTAGAGCCCTTGATAGCGTCTCTTTGTAGTGAAC	555		
Db	735	aatgaagaggttacacaatlaatglttagagcctcttgatggtlltllttttagcaatcattg	794		
QY	556	CTGAGAGGCAACGTTGTGTTTGTGTGTCAGAGATGTGACACAGTATCTAAGGTATAACCA	615		

Db 795 acagatggaagcataatatagtctgagagtgtaacttaacttaacattaca 854  
QY 616 GAAGACCTGATGAACAAGTGATATAGCATCTTGAGGACACAGCAATT 675  
Db 855 tctgactctgtgatacaagataattatccacagaagggaacattcagaggtt 914  
QY 676 GTCAAAACCT 686  
Db 915 tataaatact 925

## RESULT 8

US-09-107-847-1  
; Sequence 1, Application US/09107847  
; Patent No. 6100062  
; GENERAL INFORMATION:  
; APPLICANT: DICKNORTH, DAVID  
; APPLICANT: MICHALOVICH, DAVID  
; TITLE OF INVENTION: NOVEL USE  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Ratner & Prestia  
; STREET: P.O. Box 980  
; CITY: Valley Forge  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19482  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/107, 847  
; FILING DATE: 30-JUN-1998  
; CLASSIFICATION:  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: EP 97304996.8  
; FILING DATE: 08-JUL-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Prestia, Paul F  
; REGISTRATION NUMBER: 23,031  
; REFERENCE/DOCKET NUMBER: GH-30003  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 610-407-0700  
; TELEFAX: 610-407-0701  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 5715 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; US-09-107-847-1

Query Match 0.9%; Score 55; DB 3; Length 5715;  
Best Local Similarity 55.5%; Pred. No. 3.9e-05;  
Matches 106; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 496 AAGATGCGCTGGGCGCTATGATGCTTGAGCCCTTGATGGGCTCTTTGTAGTGAAC 555  
Db 570 AATGAAGAGTTTACACAAATATGTTAGAGCTCTTGATGTTTATTTTGAACAATCATG 629  
QY 556 CTGGAAGGCAACGTGTGTTGTGTCAGAGAATGACACAGATCTAAGCTATAACAA 615  
Db 630 ACAGATGGAGCAATATATATGTGTGAGAGTGTAACTTACTTGAACATTTTACCA 669  
QY 616 GAAGACCTGATGAACAAGTGATATAGCATCTTGAGGACGACAGCAATT 675  
Db 690 TCTGATCTTGTGATCAAAATATATTTAATTTATCCAGAGGGAACATTTAGAGGTT 749

QY 676 GTCAAAACCT 686  
Db 750 TATTAATACT 760

## RESULT 9

US-08-232-463-14/c  
; Sequence 14, Application US/08232463  
; Patent No. 5670367  
; GENERAL INFORMATION:  
; APPLICANT: DORNER, F.  
; APPLICANT: SCHEIFLINGER, F.  
; APPLICANT: FALKNER, F. G.  
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 1800 Diagonal Road, Suite 500  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22313-0299  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC Compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/232,463  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: US/07/935,313  
; FILING DATE:  
; APPLICATION NUMBER: EP 991 114 300.6  
; FILING DATE: 26-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703)836-9300  
; TELEFAX: (703)683-4109  
; TELEX: 899149  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7218 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; CLONE: PTZgpt-F1s  
; US-08-232-463-14

Query Match 0.9%; Score 53.8; DB 1; Length 7218;  
Best Local Similarity 8.1%; Pred. No. 0.0001;  
Matches 37; Conservative 225; Mismatches 197; Indels 0; Gaps 0;

QY 150 ATATGTGTTCAATATAGTGGATGGGAGAAATACCTGACCCCTCAGGCGACAGAC 209  
Db 1443 ATTTGTTACRR 1384  
QY 210 AAGAAAGCGAAGATGTCCTGACCACTTGAGCCAGCCCAAGAAAGCACTGAAA 269  
Db 1383 RRR 1324  
QY 270 ACCTAATCGTGAACAGAAATATATATAGAGAAGACTTGACAGACTGATTTTGCAAA 329  
Db 1323 RRR 1264  
QY 330 TTTAATGATATAGACACTTTTACTTCAAACTGCACAAATGTGCATCTTAAGAAAC 389

Query Match	0.83;	Score 51;	DB 1;	Length 7218;
Best Local Similarity	3.38;	Pred.No. 0.00068;		
Matches	9;	Conservative 168;	Mismatches 98;	Indels 0;
				Gaps 0;

QY 548 TAGTGAACCTGGAAGGCAACGTTCTGTT

DU 808 ICCICAGIACGAGAGAAIICACCAGCIGAIIGIIGAGAGC

Db	808	TCCTCAGTAA	CGAAGAA	TTCA	CCAC	CGATG	TTGG	AGCATT	TAGAT	GGCTTC	GCATCG	86
Qy	548	TAGTCGA	CCCGGA	AGCCAA	CGTTGT	GTTC	TGTC	AGAA	TGTG	CACAC	AGTA	60

```
Db      868 TCCTGACAAAGAGCAGCAGCATCTATGTCGACAGTATACACCTCTCTCTGGAC 927
QY      608 ATACCAAGAGAGCTGATGACAAAGTGTATATAGCATCTTGATGGGACCACA 667
Db      928 ATTACCGGGGATGTCATGATGAGAACTGTAAATTTCTTCCAGAGCAAGAACATT 987
QY      668 CGGAATTTGTCAAAAACCT 686
Db      988 CGAAGTTTATAAATCCT 1006

RESULT 12
US-08-816-693A-1
; Sequence 1, Application US/0881693A
; Patent No. 5874241
; GENERAL INFORMATION:
; APPLICANT: Takahashi, Joseph S
; APPLICANT: Turek, Fred W
; APPLICANT: Pinto, Lawrence H
; TITLE OF INVENTION: Clock Gene and Gene Product
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dressler, Rocky, Milnamow & Katz
; STREET: Two Prudential Plaza, Suite 4700
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/816,693A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5874241thrup, Thomas E
; REGISTRATION NUMBER: 33,268
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-616-5400
; TELEFAX: 312-616-5460
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7498 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 389..2954
; US-08-816-693A-1

Query Match      0.8%; Score 48.6; DB 2; Length 7498;
Best Local Similarity 53.4%; Pred. No. 0.0035;
Matches 102; Conservative 0; Mismatches 89; Indels 0; Gaps 0;
```

```
Db      887 TATAGATACT 897

RESULT 13
US-08-885-291-1
; Sequence 1, Application US/08885291A
; Patent No. 6057125
; GENERAL INFORMATION:
; APPLICANT: Takahashi, Joseph S.
; APPLICANT: Turek, Fred W.
; APPLICANT: Pinto, Lawrence H.
; TITLE OF INVENTION: CLOCK GENE AND GENE PRODUCT
; FILE REFERENCE: 0290-5
; CURRENT APPLICATION NUMBER: US/08/885,291A
; CURRENT FILING DATE: 1997-06-30
; EARLIER FILING DATE: 1997-03-13
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 7498
; TYPE: DNA
; ORGANISM: Mus musculus
; US-08-885-291-1

Query Match      0.8%; Score 48.6; DB 3; Length 7498;
Best Local Similarity 53.4%; Pred. No. 0.0035;
Matches 102; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

QY      496 AAGATGCGCTGGGGCCTATGATGCTTGAGGCCCTTGATGGTCTCTTTGTAGTGAAC 555
Db      707 aatgaagaggttacacagttacgttagagggccttgatggtttttttagcgatcatt 766
QY      556 CTGGAAGCAACGTGTGTTGTCAGAGAATGTGACACAGTATCTAAGTATAACCA 615
Db      767 acagatggaagtaataatagatctgagagtgtaactctgcattgacattacca 826
QY      616 GAAGAGCTGATGAACAAGTGTATATAGCATCTTGCACTTGCGGACACAGCAATTT 675
Db      827 tctgattcttgatcacaagtatatttattatccacagaggaacatctcagaggtt 886
QY      676 GTCAAAAACCT 686
Db      887 tataagatact 897

RESULT 14
US-09-496-672-1
; Sequence 1, Application US/09496672
; Patent No. 6291429
; GENERAL INFORMATION:
; APPLICANT: Takahashi, Joseph S.
; APPLICANT: Turek, Fred W.
; APPLICANT: Pinto, Lawrence H.
; TITLE OF INVENTION: CLOCK GENE AND GENE PRODUCT
; FILE REFERENCE: 0290-5
; CURRENT APPLICATION NUMBER: US/09/496,672
; CURRENT FILING DATE: 2000-02-03
; PRIOR APPLICATION NUMBER: 08/885,291
; PRIOR FILING DATE: 1997-06-30
; PRIOR APPLICATION NUMBER: 08/816,693
; PRIOR FILING DATE: 1997-03-13
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 7498
; TYPE: DNA
; ORGANISM: Mus musculus
; US-09-496-672-1
```



RESULT 17  
US-08-883-795A-36  
; Sequence 36, Application US/08883795A  
; Patent No. 5985607  
; GENERAL INFORMATION:  
; APPLICANT: Delcove, Genevieve  
; APPLICANT: Awang, Gregor  
; TITLE OF INVENTION: Recombinant DNA Molecules and Expression  
; TITLE OF INVENTION: Vectors for Tissue Plasminogen Activator  
; NUMBER OF SEQUENCES: 39  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BERESKIN & PARR  
; STREET: 40 King Street West  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5H 3Y2  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/883,795A  
; FILING DATE: 27-JUN-1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gravelle, Micheline  
; REGISTRATION NUMBER: 40,261  
; REFERENCE/DOCKET NUMBER: 7841-062  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 364-7311  
; TELEFAX: (416) 361-1398  
; INFORMATION FOR SEQ ID NO: 36:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 665 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
; IMMEDIATE SOURCE:  
; CLONE: Rh 32  
; US-08-883-795A-36

Query Match 0.7%; Score 45.2; DB 2; Length 665;  
Best Local Similarity 47.0%; Pred. No. 0.0061;  
Matches 140; Conservative 0; Mismatches 158; Indels 0; Gaps 0;  
QY 5796 AAACAACATATTTAAGTGGAGAACTTATGCTTTTATTTGTAATATTTTGTGAGGT 5855  
DB 3 AAAGATGAGGTAATGCTTTTATTAATTAATTTTAAATTTTAAATTTTAAATTA 62  
QY 5856 ATACATATTTGGAATTTGACTCAAAATGAGTACTTCAGTATTAATTTGATATCTCA 5915  
DB 63 AAATATTTTAAATTAATTTTATTAATTTTAAATTTTAAATTTTAAATTTTAAATTA 122  
QY 5916 TACCAATGTCCTAAAGGTGTTTGTAAAGATATCAATGCTTGATTTAGACCTAATT 5975  
DB 123 AAATATTTTAAATTAATTTTATTAATTTTAAATTTTAAATTTTAAATTTTAAATTA 182  
QY 5976 GTAGACTTAAGACTTTTATTTTCTAAACCTTGATCTGCTTATTAAGCATTTATCTA 6035  
DB 183 AAATATTTTAAATTAATTTTATTAATTTTAAATTTTAAATTTTAAATTTTAAATTA 242  
QY 6036 ATCTATATGATATGACGCCCTGTAGGAACCAATCTGTGATTTTAAATTTTAAATTT 6093  
DB 243 AAATATTTTAAATTAATTTTATTAATTTTAAATTTTAAATTTTAAATTTTAAATTA 300

RESULT 18  
US-08-785-241-2  
; Sequence 2, Application US/08785241  
; Patent No. 5655963  
; GENERAL INFORMATION:  
; APPLICANT: McKnight, Steven L.  
; APPLICANT: Russell, David W.  
; APPLICANT: Tian, Hu  
; TITLE OF INVENTION: Endothelial PAS Domain Protein  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP  
; STREET: 268 BUSH STREET, SUITE 3200  
; CITY: SAN FRANCISCO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/785,241  
; FILING DATE: 17-JAN-1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: OSMAN, RICHARD A  
; REGISTRATION NUMBER: 36,627  
; REFERENCE/DOCKET NUMBER: UTSD:1229  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 343-4341  
; TELEFAX: (415) 343-4342  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3031 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; US-08-785-241-2

Query Match 0.7%; Score 43.4; DB 1; Length 3031;  
Best Local Similarity 52.5%; Pred. No. 0.059;  
Matches 95; Conservative 0; Mismatches 86; Indels 0; Gaps 0;  
QY 515 TGATGCTTGAGGCCCTTGATGGGTCTTCTTTGTAAGCACTGGAAGCAACCTTGCT 574  
DB 368 TGACCTGGAAGCCCTTGAGGGGTTCATTCGTGTGACCAAGAGCGTGACATGATCT 427  
QY 575 TTGTGTCAGAGATGAGACACATATCTAAGTATTAACCAAGAGCGTGAACAA 634  
DB 428 TTCTGTGGGAAACATCAGCAAGTTCATGAGACTTACTCAGGTAGACTTAACGACACA 487  
QY 635 GTGTATATAGCATCTTGACATGTTGGGGACACACAGGAATTTGCAAAAACCTGTC 694  
DB 488 GCATCTTGACTTCACTCAATCCCTTGCGACCATGAAGATCCGTGAAGACCTGCTCA 547  
QY 695 A 695  
DB 548 A 548

RESULT 19  
US-08-640-906-1  
; Sequence 1, Application US/08640906B  
; Patent No. 6140100  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Gary K  
; APPLICANT: Blumenkopf, Todd A.  
; APPLICANT: Cory, Michael







```

1 MEDIUM TYPE: Floppy disk
2 COMPUTER: IBM PC compatible
3 OPERATING SYSTEM: PC-DOS/MS-DOS
4 SOFTWARE: Patentin Release #1.0, Version #1.30
5 CURRENT APPLICATION DATA:
6 APPLICATION NUMBER: US/09/056,075
7 FILING DATE:
8 CLASSIFICATION:
9 ATTORNEY/AGENT INFORMATION:
10 NAME: Seay, Nicholas J.
11 REGISTRATION NUMBER: 27386
12 REFERENCE/DOCKET NUMBER: 960296.95238
13 TELECOMMUNICATION INFORMATION:
14 TELEPHONE: 608-251-5000
15 TELEFAX: 608-251-9166
16 INFORMATION FOR SEQ ID NO: 1:
17 SEQUENCE CHARACTERISTICS:
18 LENGTH: 6243 base pairs
19 TYPE: nucleic acid
20 STRANDEDNESS: double
21 TOPOLOGY: linear
22 MOLECULE TYPE: DNA (genomic)
23 FEATURE:
24 NAME/KEY: misc_feature
25 LOCATION: 3770..4013
26 OTHER INFORMATION: /note="RP4 origin of DNA transfer (oriT) from
27 US-09-056-075-1
28 plasmid RP4"
29

```

Query Match	0.78;	Score	41.2;	DB	2;	Length	6243;
Best Local Similarly	47.58;	Pred.	No. 0.43;				
Matches 153; Conservative	0;	Mismatches	168;	Indels	1;	Gaps	1;

QY	5774	ACATAAATCCATTTATATAGCTAAACAAGCAATTAATTTAGTTGAGCACTATCGTTT	5833
Db	1362	AAATATAAAAATATATAAAATATAAAATATAAAATATATTTTATTTAAAGT	1421
QY	5834	AATGTATTAATTTTGTGAGCTATACATATGTGCAATTTGACTCAAAATGAGTACTTC	5893
Db	1422	TTGAAAAAATTTTATTTATTTATATA - ATCTTGGAGAAAGAAATATAAAAATGAGACC	1480
QY	5894	AGTATTAATATAGATATCCATAGCAATGCTCCAAAGGCTTTGTAAAGATATCA	5953
Db	1481	TTTATTAAGCCCATTTTTTTTTCATATAGCTAATATAGACGTTCTATGTTTATGTGA	1540
QY	5954	ATGCGTTGATTAAGACCTAAATTTGTAGACTTAAGACTTTTATTTCTAAACCTGTGAT	6013
Db	1541	CTTCAACATTAAGATATTTCTTTATTTTAAAGCCCTTTTCCTTAAAGGCTTTATTT	1600
QY	6014	CTGCTTATAGTATTAATCATATATATATATATATATATATATATATATATATAT	6073
Db	1601	TTTTTCTTAATACATTAAATCTCTTTTGTGTCCTTTTCCCTTTAGCTTTTAATGCT	1660
QY	6074	GATTTTATATGTTTATATCT	6095
Db	1661	CTTGATTAATTTTATTTACCTCT	1682

RESULT 25  
US-07-991-867B-8/C  
Sequence 8, Application US/07991867B  
Patent No. 5476781  
GENERAL INFORMATION:  
APPLICANT: Moyer, Richard W.  
APPLICANT: Hall, Richard L.  
APPLICANT: Gruidl, Michael E.  
TITLE OF INVENTION: No. 5476781el Entomopoxvirus Expression System  
NUMBER OF SPOUNCES: 66  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David R. Saliwanchik  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville

```

1 STATE: FL
2 COUNTRY: USA
3 ZIP: 32606
4 COMPUTER READABLE FORM:
5 MEDIUM TYPE: Floppy disk
6 COMPUTER: IBM PC compatible
7 OPERATING SYSTEM: PC-DOS/MS-DOS
8 SOFTWARE: Patent Release #1.0, Version #1.25
9 CURRENT APPLICATION DATA:
10 APPLICATION NUMBER: US/07/991,867B
11 FILING DATE: 12-DEC-1992
12 CLASSIFICATION: 435
13 PRIORITY APPLICATION DATA:
14 APPLICATION NUMBER: WO 92/14818
15 FILING DATE: 12-FEB-1992
16 PRIORITY APPLICATION DATA:
17 APPLICATION NUMBER: US 07/827,685
18 FILING DATE: 30-JAN-1992
19 PRIOR APPLICATION DATA:
20 APPLICATION NUMBER: US 07/657,584
21 FILING DATE: 19-FEB-1991
22 ATTORNEY/AGENT INFORMATION:
23 NAME: Saliwanchik, David R.
24 REGISTRATION NUMBER: 31,794
25 REFERENCE/DOCKET NUMBER: U#114.C3
26 TELECOMMUNICATION INFORMATION:
27 TELEPHONE: 904-375-8100
28 TELEFAX: 904-372-5800
29 INFORMATION FOR SEQ ID NO: 8:
30 SEQUENCE CHARACTERISTICS:
31 LENGTH: 1511 base pairs
32 TYPE: nucleic acid
33 STRANDEDNESS: double
34 TOPOLOGY: unknown
35 MOLECULE TYPE: DNA (genomic)
36 ORIGINAL SOURCE:
37 ORGANISM: Amsacta moorei entomopoxvirus
38 FEATURE:
39 NAME/KEY: CDS
40 LOCATION: complement (18..218)
41 FEATURE:
42 NAME/KEY: CDS
43 LOCATION: complement (234..782)
44 FEATURE:
45 NAME/KEY: CDS
46 LOCATION: 852..1511
47
48 US-07-991-867B-8

```

Query Match	0.78;	Score 41;	DB 1;	Length 1511;
Best Local Similarity	47.08;	Pred. No. 0.18;		
Matches 165; Conservative	0;	Mismatches 180;	Indels 6;	Gaps 1;

QY	5773	TATATTTAAATCCATTTATATATATGTAACCAACCAATTAATTTAAGTGAAGCACTATGTTT	58323
Db	1143	TACTTATATTTCTATATTTTGTATATAGATATATCTAAATATGTTATATTTTAAATTTTG	10864
QY	5833	TAAATGTATAAATTTTTGTGAGTATACATATTTGTGAATTTGACTCAAAATAGAGTACTT	58923
Db	1083	TTATATAATTTAAATTTAATATATTTTAAATTTGAATTAATATAAATCTTTAATATTTTCTG	10244
QY	5893	CAGATTTAAATTAATATCTTTCATAGCAATGTCTCTCAAAGGTGTTTGTAAAGATATTC	59523
Db	1023	GAAATATATTTTAAAAATATTTATATCATAAATATTTTAAATGCAATCTCTCAAAATTAAC	964
QY	5953	AATGCTTGATTAGACCTAATTTTGTAGACTTAAAGACTTTTATTTTCTTAAACCTGTGAT	60123
Db	963	ATTTTTTAAATATTTAATATTAATAACATATCTGTGTTATTT-----ATATTTT	910
QY	6013	TCCTGCTATTAAGCATTTATCTATATATGTATATGACACCGCTGTGAGAACCAATCTT	60723
Db	909	TTAAATATATATATTTTAAATATTTTATTTAAATTTTATACAGAAATTTAGTAATCCATTT	850

```

Oy      6073 TGAATTTTATATGTTTATATTCCTTCTTAATGAACCTTAGAAGACTACAT 6123
        |||  ||  |||  |||  ||  ||  |||  ||  |||  ||  |||
Db      849 TGAATATTTTATTTTTTTTTCATTCGATTAATTTTTTTTGAAGAAATATAT 799

```

```

RESULT 26
US-08-107-755A-8/c
Sequence 8, Application US/08107755A
Patent No. 5721352
GENERAL INFORMATION:
APPLICANT: Moyer, Richard W.
APPLICANT: Hall, Richard L.
APPLICANT: Gruidl, Michael E.
TITLE OF INVENTION: No. 5721352el Entomopoxvirus Expression System
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: David R. Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: Florida
COUNTRY: U.S.A.
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/107,755A
FILING DATE: 19-AUG-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/827,658
FILING DATE: 30-JAN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/657,584
FILING DATE: 19-FEB-1991
ATTORNEY/AGENT INFORMATION:
NAME: Saliwanchik, David R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: UFI14.C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (904) 375-8100
TELEFAX: (904) 372-5800
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 1511 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Amsacta moorei entemopoxvirus
FEATURE:
NAME/KEY: CDS
LOCATION: complement (18..218)
FEATURE:
NAME/KEY: CDS
LOCATION: complement (234..782)
FEATURE:
NAME/KEY: CDS
LOCATION: 852..1511
US-08-107-755A-8

```

Query Match ~	0.7%;	Score 41;	DB 1;	Length 1511;
Best Local Similarity	47.0%;	Pred. No. 0.18;		
Matches 165; Conservative	0;	Mismatches 180;	Indels 6;	Gaps 1

Dy 5773 TACATAAATCCCAATTATATGTAACAACAGCAATAATTACTGAGAACTTATGTGTTT 58322

Dd 1143 TACTTATATTCGCAATTTTTGTTATAGATATATCTAATATGTTATATTTTTTAAATTTTG 10844

OY	5833	TAAATGTAATAATTTTGTGGAGTACATATTCTTGGAATTCGCAATTAAGAAGACTT	5892
Db	1083	TTAATAAATTTAATTAATATATTTAAATTTGGAATATATAAAGCTTTAATTTTCTG	1024
OY	5893	CAGTATTAAATTGATATCTCCATFACCAATGTCCTCAAAGCTGTTTTGTAAGCATATC	5922
Db	1023	GAAATATTATTAAAAATATTTATTCATTAATATATATATATGCAATTTCTCTAAATTAACA	964
OY	5953	AATGCCGTGATTAGACCTAATTTTGTAGACTTAAAGACTTTTATTTTCTTAAACCCTGTGAT	6012
Db	963	ATTTTTTTAATATTAATTAATTAATTAACAATTTCTCTGTTATTT-----ATTATTTTTT	910
OY	6013	TCGCCTATATAGCATTTATCTATATCTAATATGATATGACGCCGCTGTGTGGAACCAATTC	6072
Db	909	TTAAATTAATATATTTTAAAAATTTTAAAATTTAATATATATACAGAAATTTGTAATTCATTT	850
OY	6073	TGATTTTATATATATATATATCTTCTCTATATGACCTTGAAGAAGACTCAT	6123
Db	849	TGATATATTTATTTTTTTTTCATGTGATTAATTTTTTTTTTGAAAAAATATAT	799

RESULT 27  
 US-08-544-332-8/C  
 Sequence 8, Application US/08544332  
 Patent No. 5935777  
 GENERAL INFORMATION:  
 APPLICANT: Moyer, Richard W.  
 APPLICANT: -Hall, Richard L.  
 APPLICANT: Gruidl, Michael E.  
 TITLE OF INVENTION: No. 5935777el Entomopoxvirus Expression System  
 NUMBER OF SEQUENCES: 77  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Gerard H. Bencen  
 STREET: 2421 N.W. 41st Street, Suite A-1  
 CITY: Gainesville  
 STATE: FL  
 COUNTRY: USA  
 ZIP: 32606  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patenlin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/544,332  
 FILING DATE:  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/991,867  
 FILING DATE: 07-DEC-1992  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/107,755  
 FILING DATE: 19-AUG-1993  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: WO 92/14818  
 FILING DATE: 12-FEB-1992  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/827,685  
 FILING DATE: 30-JAN-1992  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/657,584  
 FILING DATE: 19-FEB-1991  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Bencen, Gerard H.  
 REGISTRATION NUMBER: 35,746  
 REFERENCE/DOCKET NUMBER: UP114.C4  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 904-375-8100  
 TELEFAX: 904-372-5800  
 INFORMATION FOR SEQ. ID NO: 8:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1511 base pairs  
 TYPE: nucleic acid











STREET: 3054 Cornwallis Road  
CITY: Research Triangle Park  
STATE: No. 6239264th Carolina  
COUNTRY: USA  
ZIP: 27709  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/998,416  
FILING DATE: 24-DEC-1997  
CLASSIFICATION: 435  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: CH 0016/97  
FILING DATE: 31-DEC-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Meigs, J. Timothy  
REGISTRATION NUMBER: 38,241  
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976  
TELEPHONE: 919-541-8587  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 186:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 615 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: PAG1074RP  
US-08-998-416-186

Query Match 0.6%; Score 39.4; DB 4; Length 615;  
Best Local Similarity 47.0%; Pred. No. 0.28;  
Matches 162; Conservative 0; Mismatches 176; Indels 7; Gaps 1;  
QY 5776 ATAAATCCATTATATGTAACAGCAATATTTAAGTTGAGACTTANGTGTAA 5835  
DB 185 ATAAATTAATTAATTAATGAACACTATTAGTCTATGTCATTAATTTAATAGTTAT 244  
QY 5836 TTGATATATTTTGTGAGGTATACATATTTGGAATGACTCAAAATGAGTCTTCAG 5895  
DB 245 TAAATATTTATGATA-----TTATTTATTTCTTTAATAATTTATTAATGATAT 297  
QY 5896 TATTAATTTAGATATCTTCATAGCAATGCTCTCTAAAGGCTTTTGTAAAGCATATCAAT 5955  
DB 298 CAATATTAATTAATTAATTTATTTATTAATGTTTATTAATAATTAATTTATTAATAA 357  
QY 5956 GCCTTGATAGACTAATTTGTAGACTTAAGACTTTTATTTCTAAACCTTGATCT 6015  
DB 358 GATTAATTTATTTAAATATGTAATTTATTTATTTATTAATATCTATTTTAAATAA 417  
QY 6016 GCTATAGTCATTTATCTATCTATATAGATATGACGCCGTGAGCAACAATCTTGA 6075  
DB 418 TATATGCTGATTTATATTTATTTAATCTTTTAAAGAAATTTATTAATAATTAATTTAA 477  
QY 6076 TTTTATATGTTTATATCTTCTTAATGAACCTTAGAAGACTA 6120  
DB 478 CTTTAATTTCTTATTAATTTAATTTAATATTTAATAAATTATA 522

RESULT 36  
PCT-US95-06406A-11/c  
Sequence 11, Application PC/TUS9506406A  
GENERAL INFORMATION:  
APPLICANT: Janet D. Robishaw, Charles Kunsch  
TITLE OF INVENTION: CDNA Clones Encoding Human G Protein  
SUBUNITS  
NUMBER OF SEQUENCES: 23

CORRESPONDENCE ADDRESS:  
ADDRESS:  
STREET:  
CITY:  
STATE:  
COUNTRY:  
ZIP:  
COMPUTER READABLE FORM:  
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE  
COMPUTER: IBM 486  
OPERATING SYSTEM: WINDOWS FOR WORKGROUPS  
SOFTWARE: WORDPERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/06406A  
FILING DATE: Herewith  
CLASSIFICATION:  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME:  
REGISTRATION NUMBER:  
REFERENCE/DOCKET NUMBER:  
TELECOMMUNICATION INFORMATION:  
TELEPHONE:  
TELEFAX:  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 654  
TYPE: Nucleic Acid  
STRANDEDNESS: Single  
TOPOLOGY: Linear  
ANTI-SENSE: NO  
PCT-US95-06406A-11

Query Match 0.6%; Score 39.4; DB 5; Length 654;  
Best Local Similarity 56.6%; Pred. No. 0.29;  
Matches 73; Conservative 0; Mismatches 56; Indels 0; Gaps 0;  
QY 5242 TATTTTCCCTGTTTCTTTGTTAACCACGCTGACCTGATCATGCAATGA 5301  
DB 648 TTTTCTTTTCTTTTCTTTTCTTTTAAAGACAACTTATTTGAGCATCAAGCAATAT 589  
QY 5302 GAAGAAATATTTCAATTTCAATTAATGACCTTTTGTGATGACAGACTGATAC 5361  
DB 588 ACTGAAGAGTGTCATGCTCAAGTAATGTTCTTATTTCCATTAATGTACATCTGATCT 529  
QY 5362 AGCATCTCT 5370  
DB 528 CTCGCTCT 520

RESULT 37  
US-08-973-462-1  
Sequence 1, Application US/08973462B  
Patent No. 6191270  
GENERAL INFORMATION:  
APPLICANT: DRUILHE, PIERRE  
TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES  
FILE REFERENCE: 0660-0125-0 PCT  
CURRENT APPLICATION NUMBER: US/08/973,462B  
CURRENT FILING DATE: 1998-02-06  
EARLIER APPLICATION NUMBER: PCT/FR96/00894  
EARLIER FILING DATE: 1996-06-12  
EARLIER APPLICATION NUMBER: FR 95/07007  
NUMBER OF SEQ ID NOS: 29  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1  
LENGTH: 6152  
TYPE: DNA



[illegible]

```

1  APPLICANT:  Rogers IV, William O.
2  TITLE OF INVENTION:  Protective malaria sporozoite surface protein
3  TITLE OF INVENTION:  Immunogen and gene
4  NUMBER OF SEQUENCES:  2
5  CORRESPONDENCE ADDRESS:
6  ADDRESSEE:  A. David Spevack
7  STREET:  NMHC Building 1 T-12 National Naval
8  STREET:  Medical Center
9  CITY:  Bethesda
10 STATE:  MD
11 COUNTRY:  USA
12 ZIP:  20814-5044
13
14 COMPUTER READABLE FORM:
15 MEDIUM TYPE:  Floppy disk
16 COMPUTER:  IBM PC compatible
17 OPERATING SYSTEM:  PC-DOS/MS-DOS
18 SOFTWARE:  Patent Release #1.24
19
20 CURRENT APPLICATION DATA:
21 APPLICATION NUMBER:  US/07/638,431
22 FILING DATE:  19910110
23 CLASSIFICATION:  424
24 ATTORNEY/AGENT INFORMATION:
25 NAME:  Spevack, Avrom D.
26 TELECOMMUNICATION INFORMATION:
27 TELEPHONE:  (301) 295-6759
28 TELEFAX:  (301) 295-4033
29 INFORMATION FOR SEQ ID NO:  1:
30 SEQUENCE CHARACTERISTICS:
31 LENGTH:  4673 base pairs
32 TYPE:  NUCLEIC ACID
33 STRANDEDNESS:  double
34 TOPOLOGY:  linear
35 MOLECULE TYPE:  DNA (genomic)
36 HYPOTHEITICAL:  N
37 ANTI-SENSE:  N
38
39 ORIGINAL SOURCE:
40 ORGANISM:  Plasmodium yoelii
41 STRAIN:  17X(NL)
42 DEVELOPMENTAL STAGE:  erythrocytic stage
43 TISSUE TYPE:  Blood
44 CELL TYPE:  erythrocytic stage
45 IMMEDIATE SOURCE:
46 LIBRARY:  Py-lambdagt11-2-7 kb genomic expression
47 CLONE:  Py10.1111
48
49 FEATURE:
50 NAME/KEY:  CDS
51 LOCATION:  718..3195
52 OTHER INFORMATION:
53
54 US-07-638-431-1

```

Query Match	0.6%	Score 38.2;	DB 1,	length 4673;
Best Local Similarity	44.4%	Pred. No. 2,6;		
Matches	195,	Conservative	0;	Mismatches 243; Indels 1; Gaps 1
QY	5676	TTTAATTGATRGACGATATACGACAGATTTAACTCTCCATGTGTTTTATTTGCT	5735	
Db	4195	TTAAATTTAAATTTATTTTATTTAAGAACCAATATCATTTTCCAAATTTATTCATTTCT	4254	
QY	5736	TTTTTAGCAGT-GCTGACTAAGCCCAAGTTTTGTAAAGTACATAAATCCAAATTTATATG	5794	
Db	4255	TCATATTTGTAGTTGACATGCTGCATTTTTTTGTTGGAGCATATTAATTTATTTCT	4314	
QY	5795	TAAACACCAATTAATTTAAGTTGACAACCTATGCTGTTTAATTTGATATATTTTGGGAGG	5854	
Db	4315	TTATTTGCTTAATTCCTTTATTTTGTAAATTTATTAATTTTGACATATGTATATATATTTCA	4374	
QY	5855	TATACATATTTGTGAATTGACTCAAAAATGAGAGTACTTCAGTATTAATTTAGATATCTTC	5914	
Db	4375	TCCATTTTGTCTGCTTCATATTTATTAGAAAGAAATTAATTTCTATTTTGTATCAATA	4434	
QY	5915	ATAGCAATGCTCCAAAGAGCTTTTGTAAAGATATCAATGCTTGATTAACCCAAT	5974	

Db 4435 CATTATTTTATTTTCCTTTATATAAAATCGAATAAAATTCATAATACCCATTACC 4494  
QY 5975 TGTAGACTTAAGACTTTTATTTCTAAACCTTGATTCGTGCTATAAGTCATTATCG 6034  
Db 4495 TCTACCTTTTCATAATTTGAGATCATATAATGATAATAATTTATTAATGTCCCTCT 4554  
QY 6035 AATCTATATGATATGCGCCGCTGTAGAACCAATTCCTGATTTTATATGTTTATATTC 6094  
Db 4555 AGTTCGTTTATTTACATATTCCTTCCATATTTGATATTAACCAATCATGTTTCGTTT 4614  
QY 6095 TTTCTTAATGAACCTTGA 6113  
Db 4615 GTTCTTATTTAICTTATA 4633

Search completed: September 12, 2002, 09:42:11  
Job time: 21440 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 12, 2002, 03:56:26 : Search time 867.35 Seconds  
(without alignments)  
12185.767 Million cell updates/sec

Title: US-09-842-256-1  
Perfect score: 6156  
Sequence: 1 GCGGCGCCGACGCTCGGCTA.....CCACTTATGTTGTTT 6156

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

N\_Geneseq\_032802.\*  
1: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*  
2: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*  
3: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*  
4: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*  
5: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*  
6: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*  
7: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*  
8: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*  
9: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*  
10: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*  
11: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*  
12: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*  
13: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*  
14: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*  
15: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*  
16: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*  
17: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*  
18: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*  
19: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*  
20: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*  
21: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*  
22: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*  
23: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*  
24: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6152.8	99.9	6156	19	AAV03517 Human Transcription
2	597.6	9.7	4621	20	AAV99919 Murine PCIP gene (
3	547.4	8.9	4860	20	AAZ26000 AIB1 (Amplified in
4	542.8	8.8	6835	20	AAV99915 Human steroid rece
5	541.4	8.8	4789	20	AAZ80992 Human SRC-3 DNA.
6	510.6	8.3	6754	22	AAZ26480 Human cancer assoc
7	364.6	5.9	4660	21	AAO9325 Human steroid rece
8	363.2	5.9	4547	22	AAZ0356 Human immune/haema
9	189	3.1	3806	22	AAZ84490

10	187.4	3.0	3815	22	AAZ84489 Human immune/haema
11	150.6	2.4	767	22	AAH04011 Human cDNA clone (
12	150.6	2.4	1367	22	AAH17184 Human cDNA sequenc
13	106	1.7	435	22	AAH9288 Human protein encod
14	95.4	1.5	3361	20	AAZ80993 Human breast cell
15	70.6	1.1	330	22	ABA50553 Human foetal liver
16	70.6	1.1	330	22	ABA68512 Human foetal liver
17	70.6	1.1	330	22	ABA35492 Human foetal liver
18	70.6	1.1	330	22	AAK16879 Human brain expres
19	70.6	1.1	330	22	AAK42647 Human bone marrow
20	70.6	1.1	330	22	AAI23403 Human immune syste
21	70.6	1.1	330	22	AAI48724 Human immune syste
22	70.6	1.1	330	22	AAI09030 Human immune syste
23	64.8	1.1	3388	18	AAZ84543 Human steroid rece
24	58.6	1.0	2082	19	AAV41257 Human neuronal PAS
25	58	0.9	6131	24	ABL22890 Human immune syste
26	56.8	0.9	7135	22	AAZ6423 Human immune syste
27	56.4	0.9	8946	24	ABL32911 Human immune syste
28	55.6	0.9	6418	24	ABL32322 Human immune syste
29	55.6	0.9	6418	24	AAZ61073 Human gene regulat
30	55.4	0.9	2078	19	AAV41256 Human neuronal PAS
31	55	0.9	3486	21	AAZ24629 Human clock gene.
32	55	0.9	3546	19	AAV61450 Human clock gene.
33	55	0.9	5715	20	AAZ03456 Human HSCLOCK cDNA
34	55	0.9	6161	24	ABL32623 Human immune syste
35	54.2	0.9	1581	20	AAZ8984 Human transcription
36	53.8	0.9	4260	21	AAZ18177 Lung cancer associ
37	53.8	0.9	6665	22	AAZ5298 Chemically pretrea
38	53.8	0.9	6665	22	ABL32082 Human immune syste
39	53.6	0.9	15373	24	ABL22466 Human immune syste
40	53.4	0.9	6175	24	ABL33307 Human immune syste
41	53	0.9	3614	18	AAZ91883 Murine SIM (single
42	53	0.9	6065	24	ABL2504 Human immune syste
43	53	0.9	18154	24	ABL2254 Human immune syste
44	53	0.9	19087	24	ABL2793 Human immune syste
45	52.8	0.9	5947	22	AAZ4676 Tumour suppressor

#### ALIGNMENTS

RESULT 1	
AAV03517	standard; cDNA: 6156 BP.
ID	AAV03517
XX	AAV03517;
AC	20-JUL-1998 (first entry)
DT	
XX	Human transcriptional intermediary factor-2 (TIF2) cDNA.
DE	
XX	Transcriptional intermediary factor: TIF2; human; drug screening;
KW	assay; nuclear receptor; ds.
KW	
XX	
OS	Homo sapiens.
XX	
FH	key
FT	CD5
FT	Location/Qualifiers
XX	163..457
XX	/*tag= a
PN	WO9802455-A2.
XX	
PD	22-JAN-1998.
XX	
XX	11-JUL-1997; 97WO-US12100.
PF	
XX	12-JUL-1996; 96US-0021247.
PR	
XX	(BRIM ) BRISTOL-MYERS SQUIBB CO.
PA	(CNRS ) CENT NAT RECH SCI.
PA	(INRM ) INST NAT SANTE & RECH MEDICALE.
PA	(VIPA-) UNIV PASTEUR LOUIS.
XX	



Db 1561 atgacagcccccacaaagcagccctgcatgtaatccaggaagcccaactcactgctt 1620  
Qy 1621 TCACCAAGGCAATCGCATGAGCCCTTGAGAGTGGCTGGCAGCCCTCGAATGCCACCACTAG 1680  
Db 1621 tcaccaaaggcatcgcatgagccctgagtgctgagcagccctcgatcccccagctag 1680  
Qy 1681 TTTTCCCTGAGAGAGTTGCATTCCTCGGGAGTTTGAGAGCAGACAGGAATAGC 1740  
Db 1681 ttccccccgaggaagctctgcatccctctgaggagcttgcagagcagcaggaatagc 1740  
Qy 1741 CATAGTTATACCAAGCTCCCTCAATGACTTACAGCCCTCAGCGAGGGCAGCGGCTC 1800  
Db 1741 catagttatacacaagctccctcaatgacttcaagccctcagagaggcagcggtt 1800  
Qy 1801 TCATTAGGCTATGCTTGCTTGCTTACACAGACTAAAATGGCAATTTGCCAAAATCTCCCA 1860  
Db 1801 tcattagggatcgttctgcttccacagacctaaatctggccaatttgcaaaactcccca 1860  
Qy 1861 GTTAAATGAAATCTCCCTCAGTCAAGATGGGAGCTTGAGCTCAAAAGACTGTTT 1920  
Db 1861 gttaataatgaatcccccactcagcaagatgggaagcttgcagctcaaaagactgtt 1920  
Qy 1921 GGACTATATGGGAGCCCTCTGAAGGTACAACTGGACAGAGAGAGAGCTGCCATCT 1980  
Db 1921 ggaactatattgggagccctctgaaggtatacaactgagacaagagagagagcttgcact 1980  
Qy 1981 GGAAGCAAAAAGAAACAATGACCCCACTGCCCGCCGCTGAGAGAGTGAAGAGCT 2040  
Db 1981 ggaagagcaaaaagaaacaaatgaccccaactctgccccgctgtagcagtgagagagct 2040  
Qy 2041 GACGGGAGAGAGAGCTCATGAGACAGAAAGGAGACCAAACTCTGACAGTGGTGAC 2100  
Db 2041 gacgggagagagcagaactgcagacagcaaaaggcagaccaaactctgcagactctgagcc 2100  
Qy 2101 ACCAAATCTGATCAGATGAGAGCCCTGCGCTTACGACAGCTTTGTCGGAATACAAACAA 2160  
Db 2101 accaaatctgactcagatgagagccctgccttagcagactcttctgagatacaacaaa 2160  
Qy 2161 GACGCCACAGTACTGCTGCCGTGTTCTGAGGTACACACTGGAAGACTCGCTCAAGAGAG 2220  
Db 2161 gacgccacaagtgactctgagctggttctggttcaacaatgaaactctcctaaaggagag 2220  
Qy 2221 CATAAATTTTGGACAGACTCTTGACAGAGACAGATTCCCTGTGGACTTGCCCAAGTTA 2280  
Db 2221 cataaaatttgcacagactcttgcagagacagagcttcccttgcagcttgcagagctta 2280  
Qy 2281 ACAGCAGAGCCACAGGCAAAAGACTGAGCGAGAGTCCAGCAGACAGCTCTGAGATCA 2340  
Db 2281 acagcagagagccacagagcaaaagacttgcagcagagctcagcagagcagctccctgata 2340  
Qy 2341 GAAATGACTATTAAACAAGAGCCGCTGAGCCCAAGAAAGAAAGAAATGCACTACTTGGC 2400  
Db 2341 gaattgactattaaacaaagagccggtgcagcccaagaaagaaatgacactacttgc 2400  
Qy 2401 TATTTGCTAGATAAGATGATTAAGATATTGTTTACAGAGAATTAACCCCAACTT 2460  
Db 2401 tatttgcctagtaagaatgataactaaagatatgttataccaagaataatacccccaactt 2460  
Qy 2461 GAGAGACTGACAGCTAAGACAGATCTGCGCAGTAAACAAAAATTAAATGCAATGAAAACT 2520  
Db 2461 gagagactgcagcagtaagaagatccctgcagtaacacaaatataatagcaatgaaact 2520  
Qy 2521 GAGAAAGAGAGATGAGCTTTGAGCCTTGTAACCAAGCTGGCAGTGGAGACTGGACAACTTG 2580  
Db 2521 gagaagaagagagatgagcttgcagccctgggtgcacagccctgcagtgagactgc 2580  
Qy 2581 GAGAGAGATTTTGGATGATTTGCAGAAATAGTAATTAACCAAGCTTTTCCAGACAGAGG 2640  
Db 2581 gagagagatttgcagatcttgcagaaatagtaactaaccaagcttctccagacacggg 2640  
Qy 2641 CGAGGCGCCCTGCTGATCAGTTGACAAGCAAGCATCAATGACCTCATGCACTC 2700  
Db 2641 ctaggagccctgctgagctgagctgagacagcaagccatcatgaatgacatgcaactc 2700  
Db 2701 ACAGTGAAGAAACAGCCCTGTCACACACTTGTGGAGCCCAAGAAACAGCACTCGGAATTTCA 2760  
Db 2701 acagctgaagaaacagccctgtacacacttctgagagcccaagaaacagcactgcgaattca 2760  
Qy 2761 CAGAGCACTTTTAAATTAACCCAGACAGCAAGCAACTGGGACAGTTTATTTCCAAACAGANT 2820  
Db 2761 cagagcaactttaataacccacagcagaggaacttgcaggttatattgcacaaacagaat 2820  
Qy 2821 TTACCACTTGACATCACTATTCGCAAGCCCAACTGGTGGCTGAGACCTTTCCACCAATAGA 2880  
Db 2821 ttaccacttgcatacactatgcaaaagcccaactgtgtgtgcagcttccccaactcaaga 2880  
Qy 2881 AACAGTACTCCCTACTCTGATGATACCTAGCCAGGAAATGATGGGTAAATCAAGGATGATA 2940  
Db 2881 aacagtaactccctactcagtgatatacctcagcaggaatgagtggttaacaaagagatgata 2940  
Qy 2941 GGAATCCCAAGGAAATTTTGGGAGACAGTGAAGCAAGAAATGATGGTAAACAGTGCCTTCGG 3000  
Db 2941 ggaatcccaaggaatcttagggaaacagtagcacaggaatgattgtaaacagtgcttcgg 3000  
Qy 3001 CCTACTATGCACTTGAGAAATGGGACACCGAGAGATTCGAGCTGTGAGAGTCACTGACT 3060  
Db 3001 cctaactatgcactctgaggaatgggcacgcagaggttcgctgtgagagtgactaactgtgct 3060  
Qy 3061 GCTACACCAAGTCCATGTAACCGGCACTGCCAAGAGGTATGATTCGCAACCCAGCAGCC 3120  
Db 3061 gctacacaacagtgccatgtaaacccgagccagctccaagaggtatgattcggaaacccagcagcc 3120  
Qy 3121 AGCATCCCATGAGGCCCCAGCAGCAGCTGGCCCAAAACAGAGAGCTTCAGTCTCAGGTC 3180  
Db 3121 agcatcccatgagagcccaagcagcagcttgcgcaaaagacagagcttcagctcaggtc 3180  
Qy 3181 ATGAATATAGGGCCATCTGAATTAGAGATGAACATGGGGGAGCCTCACTATAGCCAACAA 3240  
Db 3181 atgaataatagggccatctgtaattagagatgaaatgagggagagccctcaglatagccaaaa 3240  
Qy 3241 CAAGCTCCTCAAAATCAACACTGCCCCAATGGCCTGAAGAACCTCGCTATAGACAGGCG 3300  
Db 3241 caagctccccaatcagaactgcagcccaatgcctggaagcactgcctatagacagagcg 3300  
Qy 3301 TCTTTTGGCAGCCAAACAGGACAGCATTTGGAGTTTCCAGATGATGACTTCTATGTTCA 3360  
Db 3301 tcttttgcagccaaacaaagcagagcatttgcaggttccagatgacttgcatagttca 3360  
Qy 3361 CATCTCGAGCTGAGTCTCCAGATGATGAGGAGAGCTCTCTGAGACAGCTGATCTGAGCC 3420  
Db 3361 catctcgagctgagctcgcagatgatatgagagagctctccctgcagcagctgtatctgccc 3420  
Qy 3421 TTGCGGAATTTTGAATGGCCTGAGAGATTTGATAGAGCCTTAGGAATACCGGAATGGTC 3480  
Db 3421 ttgcggaatTTTTGAATGGCCTGAGAGATTTGATAGAGCCTTAGGAATACCGGAATGGTC 3480  
Qy 3481 AGCCAGAGCCAAAGAGTGAATCCAGAAACAGTTTCAATCTAGAGTTCCAAATCATGCTG 3540  
Db 3481 agccagagccaaagcagtagatccagaaacagttccaaagtcaggtatccaaatactagctg 3540  
Qy 3541 GAGCAGAAAGCGCCGCTTTTCCACAGCAGTATGCAATCAAGCAAAATGGCCAGGCT 3600  
Db 3541 gagcagaaagcgccgcttctccacagcagatgcatctcagcagcaaatggcccaaggt 3600  
Qy 3601 AGCTATTCCTCCATGCAAGATCCAAACTTTTCACACCAATGGGACAGCGGCTTAAGTTATGCC 3660  
Db 3601 agctattctccatgcaagatccaaacttcaacacatgaggacagcggtcctagttatagcc 3660  
Qy 3661 AACCTCCGTATGACGCCAGACCGGGCTCAGGGCCACAGGGCTTATGTCAGAAACAGAGCA 3720  
Db 3661 aaactccgtatagcagcccaagcagcggtcctcagggcccaagggccttagtgcaaacagca 3720  
Qy 3721 AATCAACTTAAGACTTCAACTTCAGATCGCTCCAGACAGCAAGAAATGGCAGGCACTT 3780  
Db 3721 aatcaactaagaacttcaactcagatcgctccaagagacagcaaatgcagacgcaactt 3780





Db 5941 gtaagagatcaatgacctattagacctaatttgtagacttaagacttttatttct 6000  
 Qy 6001 AAACCTTGCTATTCGCTTATTAAGTCAATTTCTAATATATATGATAGCGCCGCTTA 6060  
 Db 6001 aaactcttgatctctgcttaagatcattatcattatcattatgacgcgcgtgta 6060  
 Qy 6061 GGAACCAATTCGATTTTATATATATATTCCTTCTTAATGACCTTGAAGAAGCTA 6120  
 Db 6061 ggaaccaattctgattttatattatattatatttcttcttaagaaacttaagaagacta 6120  
 Qy 6121 CATGTTACTAGACGAGCCACTTTATGGTGTGTTT 6156  
 Db 6121 catgtactaagcagccactttatgtgttttt 6156  
 RESULT 2  
 AAV9919 standard; cDNA; 4621 BP.  
 XX AAV9919;  
 AC AAV9919;  
 XX 12-MAY-1999 (first entry)  
 DE Murine PCIP gene (ortholog of human AIB1 gene).  
 XX AIB1: amplified in breast cancer; cancer; steroid: receptor;  
 KM coactivator; SCR; estrogen; ER; estrogen dependent transcription;  
 KM breast cancer; lung cancer; colon cancer; prostate cancer;  
 KM melanoma; ss.  
 XX Mus musculus.  
 OS  
 FT Key Location/Qualifiers  
 FT CDS 110..4453  
 FT /\*tag= a  
 FT /product= pcip\_protein  
 XX W09857982-A2.  
 PD 23-DEC-1998.  
 XX 17-JUN-1998; 98WO-US12689.  
 PF 17-JUN-1997; 97US-0049728.  
 PR (USGO ) US GOVERNMENT.  
 PA Meltzer P, Trent JM;  
 XX WPI; 1999-080946/07.  
 DR P-PSDB; AAMB1029.  
 XX  
 PT New isolated steroid receptor co-activator, AIB1 - used to develop  
 PT products for the diagnosis and treatment of steroid-responsive  
 PT tumours, e.g. breast, lung, prostate or colon cancers or melanomas  
 XX  
 PS Disclosure; Page 35-39; 57pp; English.  
 XX  
 CC The AIB1 protein is a member of the steroid receptor coactivator-1  
 CC (SCR-1) family of nuclear receptor co-activators that interact with  
 CC estrogen receptors (ER) to enhance ER-dependent transcription.  
 CC The AIB1 gene is amplified and over-expressed in certain cancers  
 CC in particular breast cancer and steroid hormone responsive cancers.  
 CC The AIB1 polypeptide can be used to identify compounds which inhibit  
 CC ER-dependent transcription. Increased expression of the AIB1 gene  
 CC indicates aberrantly proliferating cells, thus detection of  
 CC increased expression of the AIB1 gene or an increase in the number  
 CC of copies of the AIB1 gene can be used to diagnose cancer or a  
 CC predisposition towards developing cancer. Compounds which inhibit  
 CC expression of AIB1 or compounds which inhibit interaction of AIB1  
 CC with steroid receptors or nuclear co-factors can be used for  
 CC reducing the proliferation of cancer cells. This sequence is of the

CC murine pcip gene, a mouse ortholog of AIB1.  
 XX  
 SO Sequence 4621 BP; 1227 A; 1244 C; 1194 G; 956 T; 0 other;  
 Query Match 9.7%; Score 597.6; DB 20; Length 4621;  
 Best Local Similarity 51.4%; Pred. No. 2,36-155;  
 Matches 2199; Conservative 0; Mismatches 1864; Indels 213; Gaps 27;  
 Qy 141 CAGTTGCTGATATGTTTCAAGATGAGTGGAGAGAAATACCTTGACCCCTCCAG 200  
 Db 88 cagtgcgcagctctgacagagatgagctgagctagcgaagcctcttgatccgcgtgc 147  
 Qy 201 GGCAGAGACAGAAAGGCGAAGATGCTCCGACCAACTTGACCCGACCCCAAGAGAA 260  
 Db 148 cgtcgcagctcgcgaagcgaactgcctcgtgacgcgcgcgcgcgcgcgcgcgcgcgc 207  
 Qy 261 CACTGAAAGACGTATGCTGAGAGAGAAATATATATATATATATATATATATATATAT 320  
 Db 208 tggtagagagtgagcagc 267  
 Qy 321 TTTTGCAAAATTTTAAAT 380  
 Db 268 ctgcgcaatctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 327  
 Qy 381 AAAAGAAACGTGTGAGAGCAAAATTCGTCAGATCAAGAAAGCAAGAGAGAGAGAGAG 440  
 Db 328 aaagagagcagctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 384  
 Qy 441 CATGATGAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 500  
 Db 385 tgaatgagatgctcaaaaagctgctgctcctcagcagcagcagcagcagcagcagcag 444  
 Qy 501 TGCGCTGGGCGCTATGATGCTTGAGCCCTTGATGGGCTTCTTCTTGTGATGAACCTGA 560  
 Db 445 ctcttagagcagcttcttactacagcagcagcagcagcagcagcagcagcagcagcag 504  
 Qy 561 AGGCAAGTGTGTTGTGTCAGAGAAATGTGACAGATCAAGTATAGATATACCAAGAA 620  
 Db 505 tggaaacatgctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 564  
 Qy 621 GCTGATGAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 680  
 Db 565 cctgttaacacaaagtgtctacacagcagcagcagcagcagcagcagcagcagcagcag 624  
 Qy 681 AAACCTGCTGCCAAAGTCTATATATATATATATATATATATATATATATATATATAT 740  
 Db 625 acac---ttaccaaatccacagcagcagcagcagcagcagcagcagcagcagcagcag 681  
 Qy 741 GAACAGCATACCTTCAATTTGCGATGCTGGTAAACCTTATACCTGATTAAGAAGAGA 800  
 Db 682 aaaaagccatcatcattatgctgctgctgctgctgctgctgctgctgctgctgctgct 741  
 Qy 801 GGGTCATGATTAACAGAGAGCTCATCAAGAAATATGAACATATGAGAGTCTCGTGTCTC 860  
 Db 742 ga-----atgcagctccgaaacacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 798  
 Qy 861 TCAACCAAGTCCATCAAG 920  
 Db 799 tcagcctcgc 858  
 Qy 921 AAGAAGATTCCTCAAG 980  
 Db 859 tcgc 909  
 Qy 981 CCAGATCTCCCAAG 1040  
 Db 910 acatgaccttcgc 969  
 Qy 1041 ACAAGCTGGAG 1100  
 Db 970 gccctgcttgaagagacatacgcgaagatgtatccagagagcttccagctcagatgtg 1029

QY	1101	AGATCTGTGCTCTAATGCTTAAGAGCATCATCATGAACTGACTGAGACAGATTTGGCATTT	1160
Db	1030	gacgatactgctccag---aagctcaactataagaagcttatgtctatgctccagcaga	1086
QY	1161	CAGTCAATCTATGTTTTCCTTGCTGATGGCACTCTTGTTGGTGCACAAAGGAAG	1220
Db	1087	gaccccgatgatactgctctctctctgctgatactatcttgagtcgcagacaaaag	1146
QY	1221	CAAACTCATCCGTTCTCAGACTACTATATACCTCAACTTGTAATTTTACATATGCT	1280
Db	1147	caaatcttcgcgaactcctgtaacgagtatgctcaacgcttcactctgcaccacttct	1206
QY	1281	TCACAGAGAGACGAATGTGTGTGTGATGATCCGATCTGACTGACACAAAGATGGGAA	1340
Db	1207	tcagagaaagaagaatgatacagacaaacccaatcccgagagacaaggaactcgacc	1266
QY	1341	GCCACTGATTCAAATTAGCTCTAACGCCCTCGCCATCGGCCCTGTGCAGTGGGAACC	1400
Db	1267	tcctgcagcaaggtgtgtgcgtgacatgctccaaatcaga---tgtaagatgtagg	1322
QY	1401	AGGTGAGACATGACCCCTCATGTAGCAATATTAATTTTCCCAATAATGGCCCCAAGSAA	1460
Db	1323	gcagccgcgaactatgagtgctgcagaccaccagacacagggagatggtgagctagtc	1382
QY	1461	AATGGCGATGCCCAATGGGCAAGTTTGGTGGTTCTGGGGGATGAAACCATGTGTACGGCAT	1520
Db	1383	aacgggctctctagtagctgagctctcaactgacgcagagacaagactcaagctgcacatt	1442
QY	1521	GCACAGCAACACTCCTTCAGGGGTAGTAACATAATGACCTCAAAATGACACGCCCTCACAAAG	1580
Db	1443	cccta-----tcagaacagcagctatggtctcagcaatgacgtatcccccacagg	1491
QY	1581	CAGCCCTGGCATGTGAATCCAGAGACGCCACCTTCATGCTTTACCAAGGATCCGATGAG	1640
Db	1492	cagctccgtgctctgtgttcccaaccagcagaacatcatgattctccctcgtgaaatcg-----	1545
QY	1641	CCCTGGAGTGGCTGGCAGCCCTCGAATCCACCACAGTCAAGTTTCCCTGCAGAAAGCTT	1700
Db	1546	-----tggcccaacaaagatgagctcccaacagctctctctcgtgcagtgyc	1593
QY	1701	GCATTCCTCTGTGGGAGCTTTGACGAGCAGCAGGAATATACCCATATGTTATACCAACAGCTC	1760
Db	1594	aaacacaccatgtagactctctgcaacaacaag---gagccacagctctctctatgaagctc	1650
QY	1761	CCTCATACCTATTAGGCGCTCTCAGACGAGGGGCGACGGGCTCATATGAGGCTATCGTTGGC	1820
Db	1651	cctcagtgctcttgtaagcagcatcagtagaaggcgtgggagactctctttatctactcgtc	1710
QY	1821	TTTACCAGACCTTAATAATGGGCAATTTGCAAAACTCCCATTTAATATGATCTCTCCCC	1880
Db	1711	ctcacagagccccaactgataatctcc-----aatatgaatataaagca	1758
QY	1881	ACTGCAAGATGGGAGGCTTTGGACCTCAAAAGACTTTTGGACATATATGGGAGACCTC	1940
Db	1799	gccaagtaagtgagtgctcagagactcttaagagcccccatactctatctcgtgaacagaa	1818
QY	1941	TGAAGTACAATCTGAGACAGACAGAGACAGCTCCATCTCTGAGAGACAAAGAGAAACAA	2000
Db	1819	tcacgt-----ggagagtgtaagtggtgtcagtaacaacagcagagatcaaccacaatgaa	1870
QY	2001	TGACCCCAACTGCCCCCGCCGCTGAGACAGTGAAGAGCTGACGGCGAGACGACTGCA	2060
Db	1871	aaagaaagcaagagagagcagtgaggagtggtctcagaga-----cgcccagggagcctct	1923
QY	2061	TGACAGCAAAAGGGGAGACCAAACTCCCTGGACGCTGCTGACACCAAAATCTGATCAATGGA	2120
Db	1924	ggaaagcaaaaggccacaagaactgctgcagttaactcaagtgctcctccgcagacagag	1983
QY	2121	GCCTCGCCGCTTTACCACTCTTTGTGCGATCAACAAAGA-----	2162
Db	1984	ccattctctcttgcaacactctccctctgataccaacatgcgaagactcttcggtatgagt	2043
QY	2163	-----CTCACAGGTAGCTTGCTCGTGTGGGTCTAC	2195

Db	2044	caccagcccccctgagtgctccctccaaatcaatgaagagtgctcccaactccaagt	2103
Oy	2196	ACATGAAACCTCGCTCAAGAGCAAGACATAAATTTTGGACACAGCTTTGGACGACGACG	2255
Db	2104	gcatggtgctctgtgtgcaagaagaacaacccgatttgcacaagtgtcgcgaatgtgaa	2153
Oy	2256	TTCCCTGTGGACTTGGCCAAATTAAACGACAGACGCCACAGCCAAAGACCTGAGCCACGA	2315
Db	2164	ctccccagcagagtgccgcaaatcaatcagcagcgccatctggga-----agga	2211
Oy	2316	GTCACAGACGACACACCTCCGTGATCAGAAAGCATATTAACAAGAGCCGGTGAGCCCA	2375
Db	2212	cacgagcagcctgtcctcgttgagagaagggaacacccagcagcagcgtgagctccaa	2271
Oy	2376	GAACAAGAG---AATCAGCTACTTGTGCTATTTGCTAGATTAAGATATTAAGATAT	2432
Db	2272	gaagaagagagaataatgctctgtctgttaatactgctcgcgcagggatgtaaccagtgatg	2331
Oy	2433	TGGTTTACGAGAAATTAACCCCAACTTGAGAGACTGGACAGTAA--GACAGATCTGC	2489
Db	2332	gcttcgcaagaagctcgcagcccaagccgacagtgaggacagtaactcgaatcga	2391
Oy	2430	CAGTACACCAAAATTAATACCAATGAAACTGGAAGAGAGATGAGCTTTGACCTCG	2549
Db	2392	ctgtcccaaccaatcccagctctgycgaagaagaagccccaaatlaagaccgagacgaa	2451
Oy	2550	TGACAGAGCTGGCAGTAGCTGGACAACTTGGAGAGATTTTGGATTTTTCACAGATAG	2609
Db	2452	cgaagaagatctcggagaaaccttgataactagatgcattcttgagatttgcacagctc	2511
Oy	2610	TCAATTACACAGCTTTTCCACAGACAGAGGCGCGCCCTGCTGATCATGTTGACAA	2669
Db	2512	tgacct-ctacaacaatcctacaatbaggcggtacaccagggccaaacagcagatgttg	2570
Oy	2670	GCAAGCATTATCATATCACTCATGCACTCACAGCTGAAAAAGCCCTGTGCACCTGT	2739
Db	2571	caggaccgagttctctcgtgtgttcgaaagtcacaag-----ccgtgcagctcgt	2619
Oy	2730	TGGAGCCCAAGAAACACACACTGCGAATTTTCACAGAGCACTTTTAATTAACCCACGACAG	2789
Db	2620	tcgtctccataataacccgagcggtgtctcgtgatacctgtgtctgtgtgcagctcc	2679
Oy	2790	GCAACTGGGAGGTTATTGGCCAAACCGAATTTACCACTTGACATCACATTGCAAAAGCC	2849
Db	2680	gccagtgagaagatgctcagtgcttccctcgtgtaccaaaacgcccatactgctcgaa	2739
Oy	2850	AAGTGTCTGGACCTTCCACCAATCAGAAACAGTACTCCCTACTCGATGATACCTCA	2909
Db	2740	tccaagaaatgatacgtcagaagaattacggtgcacaatcgtggcccaacgaagaatgt	2799
Oy	2910	GCCAGAGATGATGGGTATCATCAAGGATGATAGAGAAACCAAGAAATTTAGGACAGTAG	2969
Db	2800	tcctgtgatactgcagctccctcccccggagagctggttgtaactaactcaagggccagag	2859
Oy	2970	CACAGAAATGATTGGTAACAGTCTCTCG-----GCCCTACTATGCCATCTGGAG	3019
Db	2860	aatgagcctctgcaatcaatccctcgtgaaagaaatcgtgacccgattacagcgcaatt	2919
Oy	3020	AATGGGACCCGACAGATTTGGGCTGTGAGATCACTGTGCTGTACCAACATGCCATGA	3079
Db	2920	accceagacctgcatactgggggctctgtgccttaactgtcactcgttctaactgcagctcc	2979
Oy	3080	ACCGGCTCAGTGGAGAGATGATGTTGGTAACCCACAGCACAGCAATCCCATGAGGCCCA	3139
Db	2980	aggtgcagaagaccatcgttgcgcaacagcagcagcaacagcagcagcaacaacaaca	3039
Oy	3140	GCAGC-----CAGCTGGCCAAAGACAGACGCTTCACTGCTCAGCTCATGAATATAGGCC	3194
Db	3040	gcaagcaacagcagcaacagcagcagcagcaacagcagcagatgtcttcaaatgagac	3099
Oy	3195	ATTCGATTTAGATGAACATGGGGGAGCTTCAGTATAGCCAAACAAGCTCTTCCAA	3254

```

Db 3100 tggtagatcccatbgygaatgagatcccatagc---ccagcagtcgcgtctaa 3156
QY 3255 TCAGACTGCCCATGAGCCTGAAAGCATCCTGATTAGACCGAGGCGCTTTGGACGCA 3314
Db 3157 ccaacaggttcctcggccagagggcagatctctctatgaaacaagctctcaaggtctca 3216
QY 3315 AAACAGCAGGCAATTTGGCAGTTCTCCAGATGACTTGGCTATGTCACATCTGCAGCTGA 3374
Db 3217 aaatagagcctctcttgaagaactctctgcatctgcttggtggcgaactcttaagcaga 3276
QY 3375 GTCTCCAGTATATAGGAGAGCTCTCTGACAGCTGTATCTGGCTTTGCCGAATTTTGA 3434
Db 3277 gggccagagtgacgagagagctctctgaccagctgcacacactccctgagcaacacaga 3336
QY 3435 T-----GGCTGGAGGAGATTGATAGAGCCTTAGGAATACCCGAATCGGACGACCAAG 3488
Db 3337 tggcaacaggtctcggagagagacgcagaaagggccttggaaatctcctgagctcgatcaagg 3396
QY 3489 CCAAGCAGTATAGTCCAGAAAG-----TTCTCAAGTCAGAGATTCCAAATCATATGCTTGA 3542
Db 3397 acaagcttggagtcacaacagagatgttccaaaggccaagaagcagagtaatgagtga 3456
QY 3543 GCAGAGGCGCCGCTTTTCCACAGACAGTATGATCTAGGCAACAATGGCCAGGGTAG 3602
Db 3457 tcagaagagctgcactatatagagacacataccagctcaaggtctcctccctcaagagag 3516
QY 3603 CTATTCTCCCATGCAAGATCCAAACTTTACACAC-----ATGGGACAGCGGCGCTAGTTA 3656
Db 3517 cttaaaccttaagagagcagtcacacatcglttaacctatgagtgagtcagattagcaga 3576
QY 3657 TGCCACATCCGTTATGACAGCCAG---ACCGGCGCTCAGGCGCCAGGCGCTTAGTGCAAA 3713
Db 3577 agagcagcttctctctgcaaggcagatgcatctcagagccggcctctgtagccaagaccaa 3636
QY 3714 CCAGCCAAATCACTAAGACTTCAACTTCACAGATCGCCTCCAGACAGCAAGAAATCGCCA 3773
Db 3637 caccgcgaagcagctgagatgagctcttccagtagcagaggtacagag-----ggca 3684
QY 3774 GCCATTATGAAATCAACATTCACCAATGTTTCCAAATGTGAACCTTGACTGAGCGCTGAGAT 3833
Db 3685 gcaagttttaaatacagagcgagcagcagcacttgaaatgaaatgagaacccctgcggagac 3744
QY 3834 ACCAACACAGGCACTTAAATGACAGAGTGTGGCCGACAGAGAGAGAAATCTCTGAA 3893
Db 3745 tgcgtgtagtggagccatgagcccgagcttctttaaigtccaataagtgtagccagca 3804
QY 3894 CCAGCATCTTGCACAGACAAATGATCAGCAACAGCAAGTTGACCAAGCAACTTTGAT 3953
Db 3805 gaacgagagctgagtagccatcacctgcagcagcagagagatggtgagatgagtcaca 3864
QY 3954 GATGAGAGGACAAAGGTTGAATATGACCAAGCATGGTGGCTCTAGTGGTATGCCAGC 4013
Db 3865 accaagcctcaagccttcagcccaaccccaacgtaacgcgctcccccacagagagcg 3924
QY 4014 AACTATGAGCAACCCCTGGATTTCCTCCAGGCAAAATGACAGCAAGTTCATTCTCTCAA 4073
Db 3925 ggttttggcaggttcagcaatgagcgaagccctccacaagaagtttccataltcaaga 3984
QY 4074 CTAGGAAATAGTACGACACTGATCCAGGCTTACTAGGGCTACGACTCCCGACAGGCC 4133
Db 3985 tlaagaaatggagaaacacacagagcagccttgtgtgagagctcgagctctccagcgc 4044
QY 4134 ACTTATGTCAACCCGAATGGCATATACAGAGTCCCATGATGATGCAACAGTGTCAAGGCCAA 4193
Db 4045 aatgagtcatcaagaatggaggtcttcccaagaatgcatggtgtagcaatctccctccac 4104
QY 4194 CCCAGCTATACAGGCCCTTCGACATTAATGATGGCGGAGGGAACATGGCGGGA 4253
Db 4105 acccatgatcag---ccttcagatatgaaggggtgagcgcaggaacccctgcgcagga 4161
QY 4254 CAGCATGTTTCCACAGAGTCCCAACACACTTTGGGAGAGCAACCAACACAGCATGTA 4313
Db 4162 tggctccttccccaagcagaggttggctccccaagggaacccctgcagccttacaacatg 4221

```

```

QY 4314 CAGTACACATGAGAC 4329
Db 4222 gcatatgaacagcagc 4237

RESULT 3
AAx2600
ID AAx2600 standard; DNA: 4860 BP.
XX
AC AAx2600:
XX
DE 08-JUN-1999 (first entry)
XX
DE Murine p/CIP gene.
XX
KW Mouse; p/CIP: p300/CBP/co-integrator-associated protein; gene expression;
KW breast cancer; inflammatory disease; atherosclerosis; osteoporosis; ds.
XX
OS Mus sp.
XX
FN W09856806-A1.
XX
PD 17-DEC-1998.
XX
PE 12-JUN-1998; 98WO-US1263.
XX
PR 12-JUN-1997; 97US-0049452.
XX
PA (REGC ) UNIV CALIFORNIA.
PI Glass CK, Rose DW, Rosenfield MG, Torchia J;
XX
DR WP1: 1999-080883/07.
XX
DR P-PSDB: AAM99481.
XX
PT New nucleic acids encoding p/CIP and NcoA-2 polypeptides - are used
PT to identify agents that regulate gene expression, e.g. for treatment
PT of cancer, inflammatory disease and osteoporosis
XX
PS
XX
XX Claim 3; Fig 1; 100pp; English.
XX
CC This sequence represents the nucleic acid encoding the mouse p/CIP
CC (p300/CBP/co-integrator-associated protein) polypeptide. The protein
CC can regulate gene expression so are potentially useful therapeutically,
CC e.g. against (breast) cancer, inflammatory disease (e.g. atherosclerosis)
CC or osteoporosis.
XX
SQ Sequence 4860 BP; 1274 A; 1302 C; 1251 G; 1031 T; 2 other;

Query Match 8.9%; Score 547.4; DB 20; Length 4860;
Best Local Similarity 51.1%; Pred. No. 2.5e-141;
Matches 2184; Conservative 0; Mismatches 1873; Indels 214; Gaps 30;

QY 141 CAGTTGCGATGATGATGTTCAAGATGAGTGGGATGGAGAAATACCTGACCCCTCCAG 200
Db 88 cagttgtgacccgtgcatcagatgagatgagtgactagcgaaagctcttggatccgctggc 147
QY 201 GGCAGAGCAAGAAAGCGCAAGGAATGCTTGACCAACTTGGACCCAGCCCAAGAGAA 260
Db 148 cgcgtgagctcggaaacgaacatgcctctgtagtgcctccagagcagggcttgcctacag 207
QY 261 CACTGAAAACGTTAATGCTGACACAGGAAATTAATATATATAGAAACTTGCAGAGTTGAT 320
Db 208 tggtagaagtgagcagagggagcagaggaagtaataagagtgagtgtagcagagctcat 267
QY 321 TTTTGCAAAATTTTATGATATATAGACACTTTAACTTCAAACTGACAAATGTCATCTT 380
Db 268 ctctgcgaatctcagcagacatgcgaacttcaatgtaacagccagataatgtgcatcct 327
QY 381 AAAAGAAACTGTGAAGCAAAATTCGTCAAGATCAAGAAACAGAGAAAGCAGAGCTGCCAA 440

```

328 aaeagagacagtgagacagatcagcgaataaagaacaagaaa---aacatcttccag 384  
QY 441 CATAGATGAAGTGCAGAAAGTCAGATGATCTTACAGGGCAGGGTGTCAATGCACAAGA 500  
Db 385 tgatgatagatacgaataaagctgagctctcaccagagagagatcgtataaga 444  
QY 501 TGGGCTGGGGCCTATGATGCTTGAAGGCCCTTGAATGGGTTCTTGTGTACTGAACCTGGA 560  
Db 445 cctcttagagacgctcttaactacagcagatgagatggtcttcgttctggtgagatcgaga 504  
QY 561 AGCAACGTTGTGTGTGTGTGCAGAAATGTGACACAGTATCTAAAGTATTAACCAAGAAGA 620  
Db 505 tggaaacatgtatctgctgcagaaaatgtaacacagatctgcagatacaagaagagga 564  
QY 621 GCTGATGAACAAGATGTATATATAGCATCTGCATGTTGGGACCACAGCAATTTGTCAA 680  
Db 565 cccggttaacaagaatgcttaccagcatctaacat---gagccaagaagcgaagatcttc 620  
QY 681 AAACCTGTGCCAAAGTCTATATAGTAAATGGGGGATCTTGGTCTGGCGAACCTCCGAGCG 740  
Db 621 taacacttaacaaaatccacagatgagatgagttcttggaactaagagacagagaca 680  
QY 741 GAACAGCC---ATACCTTAATTTGCGATGCTGTGTAACCACTTTACCTGATTCAGAG 796  
Db 681 aaaaagcccatatcatctttaatctgtatctgtatgtaaaaac---acacgacatcttgg 737  
QY 797 AGAGAGGCTATGATTAACCAAGCATCTCATAGAAATATGAACATGACAGTCTGCTGCTG 856  
Db 738 aagaagctgaatgcagatcccgaaaagcgcagagataatgaaacaatgagtgctcttggcc 737  
QY 857 TCTCTAACCAAGATCTCATCAAGAGAGAGAGAGATTTGCAATCTGCTGTGATTTGGC 916  
Db 798 tgcctacgctcgcgtatctgtaagaagagagagacgttgcagtgatgatactgcgcg 857  
QY 917 TGGCAGAAAGAGTTCCTCATTAAGAAAGACCACTTCTTCCCTCATCAGAAAGTTTACTA 976  
Db 858 tggctcgcgcgtgagctgcg-----ccatcccatccagatcccgagagatcttacta 908  
QY 977 CTGCGCAGATCTTCAAGGCAAGATCACGTCTGTGATACCAACCATCAGACAGCAAGCA 1036  
Db 909 ccagaaatgacatcttcgcgaaaagttgtcaatatagataaactcaactcaatgacttcca 968  
QY 1037 TGAACCAAGCTGGAGAGACCTGGTAAGAAGTGTATTCAGAAAGTTCATGCGCAGCAT 1096  
Db 969 tggagccgtgcttgaagacataatccgaagatgtaaccagaggttctcaagctgagatg 1028  
QY 1097 AAGGAGAAATCTGTCTTATGTAAGAGGCATCATGTAAGTACTAGACAGAGATGG 1156  
Db 1029 atgggacatcatgltcccaag---aagcgtacatacaagaagcttatgtcatgycacg 1085  
QY 1157 CATTCAGTCAAAATCTATCGTTTTCCTTGTGTGATGGACATCTTGTGTGCACAACGA 1216  
Db 1086 cagagaccccccgtatcgttcttccttgcgttgcgtgaactatgtgagtcgagacaa 1145  
QY 1217 AGAGCAAACTCATCGTTTCAAGACTACTAATGAACCTCAACTTTGATATCTTTACATA 1276  
Db 1146 aagaacaactcttcgcgaatccgttaacgaatgactgtaacggtcatctcgaaccact 1205  
QY 1277 TGCTTCACAGAGACGAATGTGTGTGATGAATCCGGATCTGACGTGCAAAACGATGG 1336  
Db 1206 tctcttcgagagaaacatgagatacagaccacaacccaatcccgacagaaacagagctcc 1365  
QY 1337 GGAAGCCACTGAAATTCAGTCTTAAGCCCTGCGCCCTCAGGCCCCCTGAGAGTGGGA 1396  
Db 1266 gacctcctcgacagcaggtgtgagcgtgagcatgtctccaatcagaa---tgtacaagatg 1321  
QY 1397 ACCCAGGTGCAGACATGACCTCATGATGCAATATTAATTTTCCCATTAATGGCCAAAGG 1456  
Db 1322 atgggacagccgagacataatgagtcgagacagaccacagacagcagatgaggttggagct 1381  
QY 1457 AACAAATGGGATGCCATGGCAGGTTTGGTGGTTCTGGGGGANGAACAATGTGTGAC 1516  
Db 1382 aggtacggggtctcttagtagcgtacactgaagcagagacaaagcctaacagtcgcca 1441

QY 1517 GCATGCAACCAACCACTCTCTCAGAGGTAGTAACTATGCACTCAAAATGAACGCCCTCAC 1576  
Db 1442 tctctcta-----tcgaacagcagctatgtggtctcagatgagacagatctccccc 1490  
QY 1577 AAAGCAGCCCTGCGATGAATTCACAGACAGCCCACTCCATGCTTTTCACCAAGCATGCGA 1636  
Db 1491 acgagcagctctgtctcttggcccaacagagaaacataatgattctccctcgaaacg-- 1548  
QY 1637 TGAGCCCTGGAGTGGCTGGCAGCCCTCGAATCCCACTCCAGTCAAGTTCCTCCGCGAGA 1696  
Db 1549 -----tgagcagcccaagaatgagctctccacagatctctcgtcgcag 1592  
QY 1697 GCTTGCATTCCTCTGTGGAGTGTTCAGCAGCAGCAGAAATAGCCATGATTATCCACA 1756  
Db 1593 gtgcacaatccacccaatgagacatctctgtgaaacaag---gagccaagcttctcagaa 1649  
QY 1757 GCTCCTCAATGCACTTCAGAGCCCTTCAGCGAGGGGCAAGGGGCTTCATTAGGGTATCGT 1816  
Db 1650 gctccctcagtgctctgcaagccatcagtgaaagcgttggagacctctcttatactac 1709  
QY 1817 TGGCTTCACCAAGCTTAAATGGGCAATTTGCAAAATCCCACTTAATATGAATCTC 1876  
Db 1710 tgcctccaccagccccaatcagatataatctccc-----aataagaataaa 1757  
QY 1877 CCCCACTCAGCAAGATGGGAGGCTTGGACTCAAAAGACTGTTTGGACTATATGGGAGC 1936  
Db 1758 gccagccaagtaagtgagtggtgcaggaacthaagccccctaggtctatactgtgaac 1817  
QY 1937 CCTCTGAAGGTACACTGTGCACAAAGCAGAC---AGCAGCTGCCATCTCGAGAGCAAAAG 1993  
Db 1818 agaatccatgagagatcagtggtgcagtaacaacagcagagatctcccaatgtaaaaag 1877  
QY 1994 AAACAATGACCCCAACCTGCCCCCGCGGTGAGAGTGAAGAGCTGACGGGCAAGCA 2053  
Db 1878 aagaacaagagagc-----agtggagagtgtaagaagcccaag 1919  
QY 2054 GACTGATGACCAAGAGGAGGAGACCAAACTCCTGACGCTGTGACACCAATGTGATC 2113  
Db 1920 gacctctgaaagcaagcagcacaagaactgtgagtaactcaatcagtgctcctcgaag 1979  
QY 2114 AGATGAGGCCCTGCCCCCTTATAGCCAGCTCTTGTGCGATCAAAACAAGACTCA----- 2167  
Db 1980 accgaagccatctctcttgacccaactctcccttgatccaactcgaagatcttcg 2039  
QY 2168 -----CAGGTAGCTTGCCTGATTCTG 2188  
Db 2040 ttatgttcaacagccctcctgagtggtctctcccaacatcagagagatglttccact 2099  
QY 2189 GGTCTACACATGGAACCTGCTCAAGAGAGAGCATAAATTTTGCACAGACTTTCAGG 2248  
Db 2100 ccaatgtgacatggtctctgttgcagagaaacacccgatttgcacaagttgtcagaa 2159  
QY 2249 ACAGCAGTTCCTGTGTGACTTGGCCAAAGTTAACACAGACAGCAAGCAAAAGACTCA 2308  
Db 2160 atggcacactccccaagcaggttcgccaagatacctcctcagagccttggaaagaaacga 2219  
QY 2309 GCCAGAGTCCAGCAGACACACGCTCTGGAATCAGAACTGATTAATAACAAGACCGGTGA 2368  
Db 2220 g-----cagcacgtctcctgtgagagagagaaacagcagagcagctga 2267  
QY 2369 GCCCAAGAGAAAGAG---AATGCATACTTGTGCTATTTGTAGATTAAGATGATATCA 2425  
Db 2268 gtccctaagaagaagagaaataatgctctgtcttgaatacctgtcgtgcagagatgacccca 2327  
QY 2426 AAGATATTGGTTTACGAATAATTAACCCCAACTTGAAGACTGACAGCTAA--GACAG 2482  
Db 2328 gtgattgtcttcgcaagaagctgcagcccaagcgcagcagtgaggacagatlaaactgagtc 2387  
QY 2483 ATCTGCGCACTAACCAAAATTAATAGCAATGAAGAGAGAGAGATAGCTTTG 2542  
Db 2388 agtgcagctgtctccaccaatcccaagctctgccaagagaaagaacccaataatlaagacg 2447



XX WO9857982-A2.  
PN  
XX  
PD 23-DEC-1998.  
XX  
PF 17-JUN-1998; 98WO-US12689.  
XX  
PR 17-JUN-1997; 97US-0045728.  
XX  
PA (USGO ) US GOVERNMENT.  
XX  
PI Meltzer P, Trent JM;  
XX  
DR WPI; 1999-080946/07.  
DR P-PSDB; AAW81025.  
XX  
PT New isolated steroid receptor co-activator, ARA1 - used to develop  
PT products for the diagnosis and treatment of steroid-responsive  
PT tumours, e.g. breast, lung, prostate or colon cancers or melanomas  
XX  
PS Claim 7; Page 25-30; 57pp; English.

CC The AIB1 protein is a member of the steroid receptor coactivator-1  
CC (SRC-1) family of nuclear receptor co-activators that interact with  
CC estrogen receptors (ER) to enhance ER-dependent transcription.  
CC The AIB1 gene is amplified and over-expressed in certain cancers  
CC in particular breast cancer and steroid hormone responsive cancers.  
CC The AIB1 polypeptide can be used to identify compounds which inhibit  
CC ER-dependent transcription. Increased expression of the AIB1 gene  
CC indicates aberrantly proliferating cells, thus detection of  
CC increased expression of the AIB1 gene or an increase in the number  
CC of copies of the AIB1 gene can be used to diagnose cancer or a  
CC predisposition towards developing cancer. Compounds which inhibit  
CC expression of AIB1 or compounds which inhibit interaction of AIB1  
CC with steroid receptors or nuclear co-factors can be used for  
CC reducing the proliferation of cancer cells.

Query Match	8.88;	Score 542.8;	DB 20;	Length 6835;
Best Local Similarity	51.58;	Pred. No. 6e-140;		
Matches 2241; Conservative	0;	Mismatches 1762;	Indels 351;	Gaps 29;

[illegible]

D	b	593	cggaaacattg	tattctg	tatccagaaatctgcacaaatacc	cgcaataataagcaagaga	652		
Q	y	621	gctgatgaacaaa	agctat	atagcatctgtgcagttggggacacac	acaggaattttgc	680		
D	b	653	ccgtgttaacaca	agtg	tttcaaatcttcaatgtgaagaacag	aaaggaattttctaa	712		
Q	y	681	aaacctgctgc	caaaagc	ctttatgaaatggggacattgtgtgtggggaaac	ctccagacg	740		
D	b	713	gaa---	ttt	acaaactc	tcagtttaatvgagtttccgaca	aatgaaacccaaga	769	
Q	y	741	gaacagccat	actctt	ccaattgttgagatgcgtgtaaaccttt	acctgat	tgcagaagga	800	
D	b	770	aaaaagcca	atacttaatt	aatgctgcatgtgtatgtgaacacac	acatatatttc	ggaaga	829	
Q	y	801	gggtcatgat	at	accgaacctatcagaattgtgaactat	gtcagtcgttcg	ctctc	860	
D	b	830	cataaacg	ccagatcc	ctgtaaaatg	cgacagata	tgaacaaatgcag	tgtcttgccttc	889
Q	y	861	tcaaccaa	aaagcc	ccatccttcaaaagagagattgttgacgtcc	ctgttgaatttgg	ctggc	920	
D	b	890	tcagccac	agagctat	gtatgtgaagaagggaagatttgc	aatctt	gtatagatcgtgtgc	949	
Q	y	921	aagaaagtt	cccat	tgaaaggaagaccagttcttccctcat	gcagaaaagtttt	actatcgc	980	
D	b	950	acgcgc	ctactac	agagagaagaacatttccatcaaac	ccctgagagttt	attatcacg	10099	
Q	y	981	ccagga	atctcca	aaagcaaaagmtc	acgtctcgtgatccacacacat	gagagccacat	10400	
D	b	1010	acatgat	cttccag	aaagtgatgtcaatagataacaa	atcaatgc	tgagatcc	10683	
Q	y	1041	accagc	gtggagag	acctggttaagaaagttat	ttcagaatttccat	ggccacacat	11000	
D	b	1070	gcttgctt	gtgaata	tataatccgaagtg	tatccagagatttt	agtctaagat	11229	
Q	y	1101	agaatct	gtctct	atgtactgaagacatctc	tgaagctgacac	aaagatttggc	11660	
D	b	1130	gcagctaca	gtgtccca---	gaagctcactc	atcaagaagttact	taatgagccatgc	11866	
Q	y	1161	cagtaaat	ctatgt	gttttccgtgtgtgaaagccatctgtt	gttgccacaaac	aaag	12200	
D	b	1187	aaacc	cagtatata	ctgattcgttgcgtgtgaacta	atagtgatgc	acagacaaaag	12466	
Q	y	1221	caaatct	atccgtt	tcagactactaactaagac	ctcaactgtg	taattcttt	12800	
D	b	1247	caaa	ctctccga	aatctccgttaacaaatgt	ctgacatgtgtc	taaccac	13060	
Q	y	1281	tcacga	gaagacaga	atgtgtgtgatgatccggat	ctgcagacaa	aaagtg	13400	
D	b	1307	tcagga	gaataacaga	atgtatata	agacaaacccaac	ctctgttgcaag	13660	
Q	y	1341	gccact	gtaatcc	aatgtgtctt	taacagaccctgc	ccatcagccctgtgc	14000	
D	b	1367	acctat	gtgc-----	tgatgtgcaacag	cttgcgtgagc	gcatgtatgt	14200	
Q	y	1401	aggtc	agacac	atgacctctcagtagcaat	ataataattttcc	catataat	14660	
D	b	1421	aggct	tacaaga	tgcgcagacagcaag	gctatgtcttg	agacacccatg	14860	
Q	y	1461	aatggc	atgtcc	ccatgt---g</				

Db 1661 -----tggagtlccaagaatagatgctcacaatcagtttctcctgtgc 1702  
QY 1695 AAGCTGCATTTCCCTGTGGAGTGTTCAGACAGACAGAAATGSCATAGTTATACCA 1754  
Db 1703 aggtgtgacatctccatgcaatctctgtgcaatactgagaa---ccacagcttccacg 1759  
QY 1755 CAGCTCCCTCAATGCATTTAGGCGCTCAGAGGAGGCGAGGGGTTCATTTAGGTCATC 1814  
Db 1760 cagctctcctcagtcctcagcagcatcagtgaggtgtgtgggtctcccttattctac 1819  
QY 1815 GTTGGCTTCAACAGACTTAAATATGGCAATTTGCCAAATCTCCAGTTAATGAATCC 1874  
Db 1820 tctgtctccacccagccccaattggaatactctccc-----aatagaatat 1867  
QY 1875 TCCGCCATCGACGAATGGGAAGCTTGATCTCAAAAGACTGTTTGGACTATATGGGA 1934  
Db 1868 taccacaaccaagtaagaatgaagcaatcagatcccaagagctctctgtgtcttattgcga 1927  
QY 1935 GCCCTTGAAGGTACAACTGGACACAGACAGAGCAGCTGCCATCTGAGAGCAAAAGA 1994  
Db 1928 ccaaatccagtgtagaggttcaatgtgtcagttcaataagcagagatcaacctcagtgaca 1987  
QY 1995 AACAAATGACCCCACTGCCCCCGCGGTGAGCATGAGAGAGCTGACGGGACAGCAG 2054  
Db 1988 agaaagtaa-----ggagagcagtggttgaagggcagagaatacaagggg 2032  
QY 2055 ACTGCATGACGACAAAGGAGACCAACACTCTGACAGCTGCTGACCAACAAATGTATCA 2114  
Db 2033 tcccttggaaagcaaaagatcaataaatactacgtacgtacttaccgttctctgtatga 2092  
QY 2115 GATGAGAGCCCTCGCCCTTACGACAGCTTTTGTGCGATACAAC---AAGACTCCACAG 2171  
Db 2093 ccgggggtatctctcctctgaccacactcccccctagatcaagtgtaagaatcttctgt 2152  
QY 2172 TACGTTGCTGTG-----TCTGG 2189  
Db 2153 tagtgtaccagccctctgagagctcctcctctacatctgagaggtatcctctacatc 2212  
QY 2190 GTCTACACATGGAACCTCGTCAAGGAGAGCATAAATTTTGGACAGACTCTTGACGA 2249  
Db 2213 caataatgcaatggtctactgttacaagaaagcacggatcttgcacagtgctgcagaa 2272  
QY 2250 CAGCAGTTTCCCTGTGAGCTTGGCAATTAAACAGCAAGAACCCACAGGACAGCTGAG 2309  
Db 2273 tgggaattccacagctgaggtagcaagatctacgtcagaagccactgaggaaagcacacag 2332  
QY 2310 CAGAGAGTCCAGCAGCAGCTCTGTGATGAGAAAGTACTTTAAACAGAGCCGCTGAG 2369  
Db 2333 -----cagataactctctgtggaaggaatggtgtcacaagcagagcagctaa 2383  
QY 2370 CCCCAAGAAAGAGAG---AATGCACTACTGCTTATTTGCTATGATTAAGTGTACTAA 2426  
Db 2384 tcttaagaagaagaagataatgcaatctctagatcctctgagacaggtatgatactcctcag 2443  
QY 2427 AGATATTGTTTACAGAAATAACCCCAAACTTGAGAGACTGACAGTA---GACAGA 2483  
Db 2444 tgtgtcactctctaagaactacagccccaagtgaaggtgtgataataaataagtgta 2503  
QY 2484 TCTGCGCAGTAACAAATAATTAATAGCAATGAAACTGACAAAGAGAGATGACCTTTGA 2543  
Db 2504 gtgcacacagctcaccacttctctagctcaagtaagaaagaaacctaataaagacaga 2553  
QY 2544 GCTGTGAGCAGCTGTGGAGTGTGAGCAACTTGAGAGAGATTTTGTGATTTGCA 2603  
Db 2564 gaaagtgaagagaggtactgtgagacttgatataatctagatgtctatctgtgtgtcagac 2633  
QY 2604 GAATAGTCAATTAACAGAGCTTTTCCAGACAGAGGCCAGCGCCCTGTGATGATCAGT 2663  
Db 2624 tagtgtcacttctaataatctccata-----tctcaaatggtagt 2666  
QY 2664 TGACAGCAAGCCATTCATGATGACTCATGACACTACAGCTGAAAAACAGCCCTGTAC 2723  
Db 2667 catctgggtgactaagcacaagtggttcaaggaactaa----- 2704

QY 2724 ACCGTTGGAGCCCAAGAAACAGCACTGCGAATTTTCAACAGAGCACTTTTATACCCAG 2783  
Db 2705 ---tctctgtggttgaagaatctcagagctgtgcagcttattctgtccctcataaacg 2761  
QY 2784 ACCAGGCAACTGGCGAGTATTGGCCAAACCAAAATTTACCACTTGACATCACATTGCA 2843  
Db 2762 agcagtgctctgtg-----a 2776  
QY 2844 AAGCCCAACTGTGCTGTGACCTTTTCCACCAATGAGAAACAGTAGTCCCTACTGATGAT 2903  
Db 2777 tagccctgttctgtgtgtcgaagctcccgtaaaaaatactcagtgcttctcccaagctt 2836  
QY 2904 ACCTCAGCCAGGAATGATGGGT---AATTCAGAGGATGATGTGGAACCAAGAAATTTAG 2960  
Db 2837 accaaagcaaccacatgt 2896  
QY 2961 GAAACAGTACGACGAATGATGTGTAACAGTCTTTCGCGCTACTATGCACTGTGAGA 3020  
Db 2897 ctcaagtaatgt 2956  
QY 3021 ATGGGCAACCGAGATTGCGCTGTGAGATCACCTGTGCTGTACACAGTSCCATGAA 3080  
Db 2957 ctgggtcttaccaaactcaaaagcgcggaatgtgaacctatgaatccaatccatggtg 3016  
QY 3081 CCGGCCAG-----TCCAAAGAGGTATGATTTGGAAACCCAGCAGC 3119  
Db 3017 aagaacagagagagatataatacttcttaccagacctgacgtgtgtgtgtgtgtgtgtgt 3076  
QY 3120 CAGCATCCCTCATGAGCGCCACAGCCAGCCAGCTGTGCGCAAGAC-----AGAGCTTCACTC 3173  
Db 3077 caactgtcctctgt 3136  
QY 3174 TCAGGTATGAAATATGAGCCCATCTGAAATTAAGATGAACATGGGGGACCTCATATAG 3223  
Db 3137 gcagatgtcttaaaatgagcgt 3196  
QY 3234 CCAACAAACAGCTCTCCCAATTCAGACTGCCCTGAGCTGAAAGCAATCCGCTATAGA 3293  
Db 3197 ccaagcagcagcatct---aaccaatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3253  
QY 3294 CAGGCGCTTTTTCAGCCCAAGACAGCCACATTTTGGCGAGTTTCTCAGATGACTTGTCT 3353  
Db 3254 acaagttctcatgtgcactcaataatagccctctctttaggattccctgtatgtctgt 3313  
QY 3354 ATGTCCATCTCTCAGCTGAGTCTCCGAGTGTGATGAGGAGCTCTCTGACACAGCTGTA 3413  
Db 3314 tgggtccacacttccaaactgtgaaggtcagaggtgacgaagaagcatltagtgaaccagctgca 3373  
QY 3414 TCTGCGCTTGGGGAATTTTGTAT-----GGCTGAGAGAGATTTGATGAGCTTATGAGAT 3467  
Db 3374 caactctctcagcaacaacagatgcccacagagcgtggaagaatgtgacagagcttgtggcat 3433  
QY 3468 ACCCGAAGCTGTGACGACAGCCAGACAGTATGATCCAGAACAG-----TTCCTCAAGTCA 3521  
Db 3434 tctctgaactgttcaatcaatcaaggtgagcatttaggcccacaaggaatgcttccaagcca 3493  
QY 3522 GGATTCCAACTATCTGCTGAGCAGAAAGCGCCGTTTTCACACAGCAATGATGATGCA 3581  
Db 3494 aagaagcagcagtaatgtatgtatcagaagcaggaattatgtgacaagacataccagaca 3553  
QY 3582 GGCACAAATGGCCAGGCTGTATTTCTCCCATGCAAGATCAAACTTTTCAACACATGGG 3641  
Db 3554 ggggtcccccacatgtcaaggaaggtcttcaatctcgaaggaacaatccatctttaaactat 3613  
QY 3642 ACAGCGGCTAGTTATGCAACACTCCGTATGCAAGCCAG-----CCGGGCTCAG 3692  
Db 3614 gatgaatcagatgaaacagcaaggaatcttctcctcacaaggaatgtacccaagagcaaa 3673  
QY 3693 GCCACGCGGCTAGTGTGAGAACCAAGCCAAATCAACTTAAGACTTCAAGCATGCGCT 3752  
Db 3674 catcatgagaccccggaacaacacccccaagcaacttagaatgtcaggtctcagcagaggt 3733









QY 2844 AAGCCCACTGGCTGGAGCTTTCCCAATCAGTAACAGTAGTCCCTACTCAGTGAT 2903  
 Db 2761 tagccctgttctgtgtgtcgaagtcctccagtaaaaaatacagtgcttcccaatgt 2820  
 QY 2904 ACCCTGACCCGAGATGATGGGT---AATCAGGATGATAGCAAAACGAATTTAG 2960  
 Db 2821 aacaaagcaaccatgtgtgtgtggaatccaagaatgtatgtatgagaataatgat 2880  
 QY 2961 GAACAGTACCAAGAAATATGTGTAACAGTGTCTGTGGCTACTGATCCCACTGAGA 3020  
 Db 2881 ctaagatgagtggtggtgcaaacgaatgtacgtgactgacccctcccaagaga 2940  
 QY 3021 ATGGCACCCGACAGATTCGGCTGTGAGACTCCTGTGTACCAACCAAGTCCATGAA 3080  
 Db 2941 cgtgggtcttacaactcaaaagccgycagatgtgaacctatgaatccaactcattg 3000  
 QY 3081 CCGGCCAG-----TCCAGAGGTATGATTCGAGACCCAGCAGC 3119  
 Db 3001 aagaccagagagatataactcttctaccagacctgactgtgtgtgtctatctcc 3060  
 QY 3120 CAGCATCCCATGAGGCCAGCAGCAGCTGGCCAAAGACA-----GACGCTTACGTC 3173  
 Db 3061 caagtgtcctctcggtccataatagcaccaggtgcgagaccaatgtgcaaacagaga 3120  
 QY 3174 TCAGGTCAATGAATATAGGSCCATCTGATTTAGATGAACATGGGGGAGCTCAGTATAG 3233  
 Db 3121 gcaagatgtcctaataatgagcctgtgtgaatcccatggtgaatgtgggtcattcccat 3180  
 QY 3234 CCACACAAAGCTCTCCAAATACAGCTCCCATGCTGAAAGCATCTCTGCTATGA 3293  
 Db 3181 ccaagcagagacatct---accacatgtgttctgtcccgatgtgactgtgtcattgga 3237  
 QY 3294 CCAGGGGTCTTTTGGCAGCCAAACAGCAGCAGCTTTGGCAGTGTCTCCAGTACTTGCT 3353  
 Db 3238 acaagttctcatgtgacccaataatagcctcttcttgaagaatccctgtgatatctgt 3297  
 QY 3354 ATGTCCACATCTCGACGTAGTCTCCAGTATGAGGAGAGCTCTCTGAGCAGCTGTA 3413  
 Db 3298 tgggccaacctccaacctggaagcagagatgacgaagaacattatgtgacagctgca 3357  
 QY 3414 TCTGGCTTGGCGAATTTTGAT-----GGCTGGAAGATTTGATAGACCTTAGAAT 3467  
 Db 3358 cactcttctcacaacacagatgtccacagcctggaagatgtgacagacttgggcat 3417  
 QY 3468 ACCCGAATGTGTACAGCAGCAGCAGAGTATGATCCAGAACAG-----TTCTCAGTCA 3521  
 Db 3418 tctgaactgttcaatcagagacagcatatgagcccaacagatgtcttccaagcca 3477  
 QY 3522 GGATTCCAATCATCATCTGTGAGAGAGGCGCCGTTTTCACACAGCATATGATCTCA 3581  
 Db 3478 agaagacagcatatgtatgataagaagatataatgtgacagataaccagacaca 3537  
 QY 3582 GGCACAAATGGCCAGGATATCTCTCCATGCAAGATCCAAACTTTACACACATGGG 3641  
 Db 3538 ggggctccaatgtcaagagagcttctcatctcagggacatccactttaaactat 3597  
 QY 3642 ACAGCGCCTAGTTATGCCACACTCCGTATGACGCCAGA-----CCGGGCTTCAG 3692  
 Db 3598 gatgaatcagatgaacacagagcaatttctctccaagatgacacccaagagacca 3657  
 QY 3693 GCCCAGGGGCTTAGTGACAGAACCCAAATCACTTAAGACTTCAACTTCAGAGCTGCT 3752  
 Db 3658 catcatgagaccccgagacaacaccccaagcaactagatgtcagcttcaagagagct 3717  
 QY 3753 CCAAGCAGCAGAGAAAT-----CGCCAGCACTTATTAATCAATCAGCATGT 3800  
 Db 3718 gaaagggccagcagtttctgaatcagacagcagcactgtgaatgaataatgaaaccc 3777  
 QY 3801 TTTCCAATG-----TGAACCTTGACTGTAGGCTCGAGCTGAGTACCAACAGCGC 3845  
 Db 3778 tactgt 3837

QY 3846 ACCATTAATGACAGATCTGGCCCA----- 3872  
 Db 3838 tttcttaatgtcctaattgttcgccaacgacagagagctgtcgaatcacttcg 3897  
 QY 3873 -----GAGACAGAGGGAATTCGTAACCCAGCATCTTCGACA 3908  
 Db 3898 acaacagaggtgtgtatgtatgtatgtatgtatgtatgtatgtatgtatgtatgtat 3957  
 QY 3909 GAGCAAAATGCACTACAGCAACAGCAAGTTCCAGCAAGCAACTTTGATGATGAGAGCAAG 3968  
 Db 3958 gcaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 4017  
 QY 3969 GTTGAATATGACACCAAGCATGTGTGTCTCTAGTGTATGCAACACTATGACAAACC 4028  
 Db 4018 cttagcccaactcttaatgtgtactgttcccccagcatgtatgtgtgtgtgtgtgtgtgt 4077  
 QY 4029 TCGGATTCGCCAGGCAATAGCAGAGTTCATTTCTCCAACTAGCAATAAGTCA 4088  
 Db 4078 caaatgtcccaagctctcccgcaacagtttccataccaacaaatltatgtatgtatgtat 4137  
 QY 4089 GCAACCTGATCCAGGCTTTACTGTGGGCTACGACTCCCGAGACCCGACTTATGTACCCCG 4148  
 Db 4138 acaaccagatccaagcttctgt 4197  
 QY 4149 AATGCACTATACAGAGATCCCATGATGACAGATGTCTAGGCCAACCCAGCTATCAGCC 4208  
 Db 4198 aatgtgtcctccccaatgccatgtatgtacacacccgagctgtgcatctatag-- 4255  
 QY 4209 CCCCTCCGACATTAATGATGGGCGCAGGGAACATGGGCGGAACAGCATGTTTCCCA 4268  
 Db 4256 -tctcagaagaatgaaaggt 4314  
 QY 4269 GAGTCCCCACCACTTTGGCAGCAGCAACACACAGCATG 4311  
 Db 4315 gaaagcagttgtccaccaggggaatctctgagtgtatgtatg 4357  
 RESULT 6  
 AAF26490  
 ID AAF26490 standard; DNA; 6754 BP.  
 XX  
 AC AAF26490;  
 XX  
 DT 27-MAR-2001 (first entry)  
 XX  
 KW Steroid receptor coactivator-3; SRC-3; antisense; infection;  
 XX inflammation; tumour; cancer; ds.  
 KW  
 OS Homo sapiens.  
 XX  
 PN US6156571-A.  
 XX  
 PD 05-DEC-2000.  
 XX  
 PF 15-NOV-1999; 99US-0440612.  
 XX  
 PR 15-NOV-1999; 99US-0440612.  
 XX  
 PA (ISIS-) ISIS PHARM INC.  
 XX  
 PI Bennett CF, Cowser LM;  
 XX  
 DR WPI; 2001-079549/09.  
 XX  
 PT Novel antisense compound useful to prevent or delay infection,  
 XX inflammation or tumor formation, specifically hybridizes with and  
 XX inhibits the expression of human steroid receptor coactivator-3  
 PS Example 13; Column 43-54; 36pp; English.  
 XX  
 CC The present invention relates to an antisense oligonucleotide.

CC targeted to a nucleic acid molecule encoding human steroid receptor  
CC coactivator-3 (SRC-3). The invention is useful for inhibiting the  
CC expression of SRC-3 in human cells or tissues *in vitro*. It is  
CC useful for diagnostics, therapeutics, prophylaxis and as  
CC research reagents and kits. It is useful prophylactically, to  
CC prevent or delay infection, inflammation or tumor formation.

**SQ** Sequence 6754 BP; 1955 A; 1463 C; 1511 G; 1825 T; 0 other;

Query Match	8.38;	Score 510.6;	DB 22;	Length 6754;
Best Local Similarity	51.18;			
Matches 2219; Conservative	0;	Pred. No. 5.8e-131;		
		Mismatches 1769;	Indels 357;	Gaps 29;

[illegible]

OY	1041	ACGAGCTGGGAGGACCTGGTAAAGAGGTATATCGAAGTTCCATGGCGACATGAAGG	1100
Db	1053	gcccgtcttgaagataataaccgaagtgatctcagagatcttcttaagcttaaaatgag	1112
OY	1101	AGAACTGTGTCCCA-----TGCCTAAGAGGACATCA	1133
Db	1113	gcagatcatggtcccaagaagaagctacatacaagaagttaccagtgatggatatttcccc	1172
OY	1134	TGAAGTACTGAGACAGGATTTGCATTGATCAATCTATGCTTTTCCCTGTCTGATGG	1193
Db	1173	aacagcttatcttaatgycatgcagaaacccagatatcatgcttcgttgcgtatg	1232
OY	1194	CACCTTTGTTCGTCACAAAGCAAGGCAAACTCATCCGCTTCACAGTACTACTAATGAAC	1253
Db	1233	aactatagttgcgtcacagacaaaagacaactcttcgaaactctgtaacaaatgacg	1292
OY	1254	TCACCTTGTATATCTTTACATATATGCTTACAGAGACGCAAGTATGTGTGTGATGAATCC	1313
Db	1293	acatggtcttgcttccaaccactcttccttaagagagacaagaatgatatagaccaaaccc	1352
OY	1314	GGATCTGACTGGACAAACGATGGGGAAAGCCACTGAAATCCATTTACTCTTAACAGCCCTGC	1373
Db	1353	aaatccgtgttgacaagaagatctagaccactatg-----gctggtatccaacgttcggt	1406
OY	1374	CCATAGGACCCCTGTGCGATGGGAGCCGATGAGCAATGATGACCTTCACAGACATATATAA	1433
Db	1407	agcgcgcatgagatctgctgcaccaaaacgaagcttacaagatgacgagacgaagggctatg	1466
OY	1434	TTTTCCATTAATGCGCCCAAGGACAAATGGGCATGCCCATG---GGCAGGTTTGGTGG	1490
Db	1467	cttgacagacccttagcacacacagggacagatgagtggagctagtagtattggygtcttcagtaa	1526
OY	1491	TTTCGGGGGAATGAAACCATGTGTACAGGATGCAAGCAACCATCTCT---CAGGGTAGTAA	1547
Db	1527	catagctctatgaccccttgagcgacatgacatcacatcttccatacsgaaacaaacaa	1566
OY	1548	CTATGCACTAAATGAACAGCCCTTCACAAAGACGCCCTGGCAATGCCAGCAAGCC	1607
Db	1587	ctatagctcaaatgagtagccccccacatgtagtctctgtctctgcaccaaacgcga	1646
OY	1608	CACCTCATGCTTTACCAAGGACGATGCGATGAGCCCTGGATGTGCTGGCAGCCCTCGAAT	1667
Db	1647	gaatatcatgattctctctcgtatgag-----tggagatccaaagat	1688
OY	1668	CCCAACCGACAGTATTTCCCTCGACAGAGCTTGATTCCTCTGGGAGATTGGACAG	1727
Db	1689	agcctcaatcagtttctctcgtgtgacggtgtagcactctccatgagcatctcttgagca	1748
OY	1728	CACAGAAATAGCCATATGTTATACCAACAGCTCCCTCAATGTCATTCAGGCCCTCAGCA	1787
Db	1749	tactggtgaa---ccacagctttccagcagctctctcagtgccctgcgaacgcatcagya	1805
OY	1788	GGGGACAGGGGTCTCATTTAGCGTCACTGTTGGCTCACACAGACTCTAAATAATGGGCATTT	1847
Db	1806	aggtgtgtgagacttcccttcttactactctgtcatcacaagggccccaattggtatactc	1865
OY	1848	GCAAAATCCCAAGTATTAATGATGATCTCTCCCACTACGACAAATGATGGGAAGCTTGACATC	1907
Db	1866	tccc-----aatatgataatacccaacccaagtaagtaagcaatcaggttc	1913
OY	1908	AAAAGACTGTTTGGACTATATGGGAGCCCTCTGAAGGTACACTGTGACAAACGAGAGAG	1967
Db	1914	caagagctcctcgtgcttattatgacccaanaatccagtytgaggttcaatgtytgagtc	1973
OY	1968	CAGCTGCATCTCGAGAGCAAAAGAAACAATGACCCCAACTGCGCCCGCGCGTGAAG	2027
Db	1974	aaatagcaagatcaactcagtgacaagaagaagtaagagagc-----ag	2018
OY	2028	CAGTGAAGAGCTGACGGGCGACGACACTGCATGACGCAAAAGGCGACAGCAAACTCTT	2087
Db	2019	tgttgaggggacagatactaaaggggtctcttggaagcaaggtctctaaaaattact	2078

QY	2088	GCACGTGCTGACCCACAAATCTGATATGAGAGAGCCCTGCGCCTTACGACAGTCTTGTGTC	2147
Db	2079	gcagcttacttaccctcgtctctctctgataccggggtcattctccctcttgaccacatctccccc	2138
QY	2148	GGATTACAA---ACAAAGACTCCACAGTAGCTTGGCCTTG------	2183
Db	2139	agattcaaagttgtaagaatactctctgttgatgcacagccccctctgagttctctctc	2198
QY	2184	-----TTTCGGGTCCTACATGGAACCTCGCTCAAGAGACGA	2222
Db	2199	tacatctgagagtatctccctacatccaatatgatgatgtaacgttataagaagaagca	2258
QY	2223	TAAAAATTTTGCACAGACTCTTTGCAGGACAGAGACTTCCCTGTGAGACTTGGCCAACTTAAC	2282
Db	2259	ccggattcttgcaagaattctgtctgagaatgagaaatlccacagcttgagtagccaatattac	2318
QY	2283	AGCAAGAACCCACAGGACCAAGACCCGTGAGGCCGAGAGTGTCCACACACACAGGCTCGTGATCGA	2342
Db	2319	tgcgaagaagccactctgggaagaacacag-----cagtaatactcttgtygggaag	2359
QY	2343	AGTGACTTATTAACAGAGCCGGGTGAGCCCCCAAGAAAGAG---AATGCATACTTTCG	2399
Db	2370	aaatgtgtcgaagcagcagcagtaagtcctaagaagaaggaataatgcaactcttag	2429
QY	2400	CTATTTGCTAGATTAAGATGATACTATAAGATATTTGGTTTATCCAGAAATTAACCCCAAACT	2459
Db	2430	atacctgtctgacagcggatgatactcctgtagtgacactctctaaagaactacacagccccaagt	2489
QY	2460	TGAGAGACTGTGACAGTAAAGACAGATCC---TGCACATTAACACAAAATTAAATAGCAATAA	2516
Db	2490	ggaagagatgtaataataatgtagtcagtcagcagctccacatctctagctcaagtca	2549
QY	2517	AACTGAGAAGGAGAGATGAGCTTTTGAGCCTGTGGACACGCTTGACAGTGAAGTGACAA	2576
Db	2550	agagaagaagccctaataatlaagacagagacaagtgaagaaggatctgagacttgataa	2609
QY	2577	CTTGAGAGAGATTTTGGATGATATTTGGCAATATATCTAATTAACACAGCTTTTCCAGAAC	2636
Db	2610	lctagatgctattcttctgtgtatcgtcagtagttcttgactttaacaatattccatctctc	2669
QY	2637	GAGGCAAGGCCCCCTCTGTGATCAGTTGATGTCAGCAAGCAAGCATCATCAATGACCTCATCA	2696
Db	2670	aaa-----tgtagtgcattcttgagggaactcaagcaacagtgltttcaagga	2712
QY	2697	ACTCACAGCTGAAAAAGCCCTGTCAACCTGTTTGGAGGCCAGAAAACAGCACTGCGAAT	2756
Db	2713	actaa-----ttctcgyglttgaaagatgtaacagtcagtcgt	2747
QY	2757	TTTCACAGAGCACTTTTAATTAACCCAGACGACGAGGCAACTGGGCAAGTATTTGCCAAACCA	2816
Db	2748	gcagctctattcgctcccatataaacagagagatgctctcg-----	2788
QY	2817	GAATTTACCACTTGACATCACTTGCAAAAGCCCAACTGGTGGTGAGCCTTTCACCAAT	2876
Db	2789	-----atagccctgtttcttgtagtcaagtctcccaagt	2822
QY	2877	CAGAAACAGTAGTCCCTACTCAGTAGATACCTCCAGCCAGGAATGATGGT---AATCAAG	2933
Db	2823	aaaaataatacagtgcttccccaatgltacaagaagcaaccaatgtytggtyggaataccaag	2882
QY	2934	GATGATAGAAACCAAGGAATTTAGGGAACAGTACACACAGGAATGATTTGGTACAGTGC	2993
Db	2883	aatgtatgtagtcaaggaataatltatgctcaagt-----atbgagagatggg	2930
QY	2994	TTTCGGGCTACTATGCAATCTGGAGAAATGGGCAACCGCAGAGATGTGGCTGTGAAGTCAAC	3053
Db	2931	cttacaacaactcaaaagccggcagaaatgaaacctatgaaatlcaaaactcaatgtygaagac	2990
QY	3054	CTGTGCTGCTACCAACAGTGCGCATGAACCGGCAGTCAACAGGAAGTATGATTCGGAGACC	3113
Db	2991	agggagagatatataactctcttaccacagactctgacgtgggttgctctattccccaatt	3050
QY	3114	AGCAAGCCGATCCCATGAGGCCACAGACCCAGCTGGGCCAAAGACAGACGCTTTCAGTC	3173

Db	3051	gccccttcggtctaataagcatlaccaggtgcgagaccagatltggcaacag-----cagca	3104
QY	3174	TCAGGTCATGAAATATAGGCCATCTCAATTTAGATGAATGAATGAGGAGACCTCATGATATG	3233
Db	3105	gcagatgtcttcaaatatgagcgcttggatgaatccccatgggaatgggggtcaatccctatg	3164
QY	3234	CCAAACAAACAGCTCTCTCAATATACAGCTGCCCATGCGCTGAAGACATCTGCTATTAGA	3293
Db	3165	ccaagcagcagcatct---aaccacttgggttccctgcccgaatggcatgtgtccatltga	3221
QY	3294	CCAGGCGCTTTTGGCCGCAAAAACAGGCGCATTTGGCAGTTCCTCCAGATGAGCTTGGT	3353
Db	3222	acaagttcttcaatgagcctcaaaaatagcgctctcttcaaggaatccctggatagatctgt	3281
QY	3354	ATGTCCACATCTTCACAGCTCAGTCTCCGATGTGAGGAGGCTCTCTGGACACAGCTGA	3413
Db	3282	tgggcacaacttccaacttgaagcgagcgatgacgaaagagcatattatggacacgctgca	3341
QY	3414	TCCTGGCCTTGGCGAATTTTGAT-----GGCTCGAGAGAGATTGATAGACCTTAGCAT	3467
Db	3342	cactcttctcagcaacacagatgtccacagcgctggaagaatltgacagagcttgggcat	3401
QY	3468	ACCCGAACTGGTTCAGCCAGAGCCAAAGATAGATCCAGAAACG-----TTCTCAAGCA	3521
Db	3402	tcttgaactgttcaatcagagacagcgatataagagcccaacagatgtcttccaagcgca	3461
QY	3522	GGATTTCCACATCATCTGCTGGAGCAGAGAGGCGCCGTTTCCACAGCATATGATCTCA	3581
Db	3462	agaagcagcagtaataatgtatgatacagaagagatataatgacagataatcccaagca	3521
QY	3582	GGCACAAATGGCCAGGGTATGCTATTTCTCCATGACAAATCCAAACTTTACACCAATGG	3641
Db	3522	ggggcctccaatgaagaagaggtcttcatcttcacaggaacatcacatctttaaactctat	3581
QY	3642	ACAGGGGCTAGTTATATCCACACTCCGTTATGAGCCCCACA-----CCGGGCTGAG	3692
Db	3582	gatgaatcagatgaacacgaagaagcaatlttctctccaaggaatgacccaagcgcca	3641
QY	3693	GCCACGGGCTTATGTCAGAACCCAGCCAAATCAACTTACACTTTCATCTTCAGCATCGCT	3752
Db	3642	catatgagaccccgagcaaacacccccaagcaacttgaatgaagcgtcttcagcagaagct	3701
QY	3753	CCAAACACAGCAGATGCGCCAGC-----CACTTATGATCAATATCAG	3794
Db	3702	gcagggccagcagatlttgaatcagagccgacagcacttgaatltgaatatgaaaaccc	3761
QY	3795	CAATGTTTCCAAATGTGAACCTTGACTCTGAGGCTTGAGTACCAACACAGGCCATATTA	3854
Db	3762	tactgtcgtgtgtgtcgtgcgtgtatgaggtctatgatgtcagcccgaggggttctctaa	3821
QY	3855	TGCACAGATGCTGGCCGACAGACAGAGGGAATCTT-----	3890
Db	3822	tgctcaaatgtgcgcccaagcgacgagagagtcgtctaagtatacttccgaacaacagag	3881
QY	3891	-----GAACCAAGCATTTTGGACAGACAAATGATCATCAGA	3926
Db	3882	ggtgtctatgtatgtacgacgacgaacacagcagcagcagcagcagcagcagcagca	3941
QY	3927	ACAGCAAGTTACGACACGAACCTTTGATGATGAGAGSACAAGGTTGATATGACACCAAG	3966
Db	3942	gcaacagaaacagaaacagaaacagaaacagcagcaaacccagccttcacgccaactctaa	4001
QY	3987	CATGGTGGCTCTATGTTGTATGGCAGCAACTATGACAACTCCGATTTCCCAAGGCAAA	4046
Db	4002	tgtactgtcttcccacagcatgtgattggcgttlttgcaggaaaccacaatgtgcccacgctc	4051
QY	4047	TGCACAGCAAGTTTCATTTCTCTCAAACTACGGAATTAATCTGACAACTGTATCCAGGCTT	4106
Db	4062	tccgcaacagtttccatatacacaacaatatgtatgggaacaaacacagatccagctg	4121
QY	4107	TACTTGGGGCTACGACTCCCAAGAGCCCACTTATGTACCCCGAATGCAATACACAGAG	4166

Db 4122 tgcgtcagtgctagctccccaatgcaatgctgcgtcaagaatgggtccctcccaaga 4181  
Qy 4167 TCCCATGATGCAACAGTCTCAGGCCAACCCAGCTATCAGGCCCTCCGACATTAATGG 4226  
Db 4182 tcccatgcatgcaaacaccgcagctgcagctcattcatcag---tccctgaagaatgaaagg 4238  
Qy 4227 ATGGGCGCAGGGGAACATGGCGGAACAGCATGTTTTCAGCAGATCCACACCACTT 4286  
Db 4239 ctggccatcaggaatttggccaggaacagctcttcccaagcaagtttgcaccca 4298  
Qy 4287 TGGGCGACGACCAAAACACGACATG 4311  
Db 4299 ggaggatccctgcagtgatagatg 4323

RESULT 7  
AAA09325  
ID AAA09325 standard; DNA: 4660 BP.  
XX  
AC AAA09325;  
XX  
DT 10-AUG-2000 (first entry)  
XX  
DE Human cancer associated antigen precursor DNA, clone NY-REN-52.  
XX  
XX renal cancer: cancer associated antigen precursor; diagnosis;  
KM cytosolic; steroid receptor coactivator; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200020587-A2.  
XX  
PD 13-Apr-2000.  
XX  
PF 04-OCT-1999; 99WO-US22873.  
XX  
PR 05-OCT-1998; 98US-0166300.  
PR 05-OCT-1998; 98US-0166350.  
XX  
PA (LUDWIG-) LUDWIG INST CANCER RES.  
PI Obata Y, Gout I, Tureci O, Sahin U, Pfreundschuh M, Scanlan MJ;  
PI Stockert E, Chen Y, Old LJ, Jager E, Knuth A;  
XX  
DR WPI; 2000-303774/26.  
XX  
XX Preventing, diagnosing and/or treating disorders associated with  
PT abnormal expression of human cancer associated antigens  
XX  
PS Claim 57; Page 96-97; 121pp; English.  
XX  
CC AAA09321-45 were isolated by SEREX screening from a renal cancer  
CC cell line 1973/10.4. Homology searching revealed that these clones  
CC correspond to known genes. The present sequence has identity with the  
CC steroid receptor coactivator gene. The genes encode cancer associated  
CC antigen precursors. These products are useful in methods for  
CC preventing, diagnosing and/or treating disorders, especially cancer,  
CC associated with abnormal expression of human cancer associated antigens.  
CC The method comprises contacting a sample from a subject with an agent  
CC that specifically binds to the nucleic acid molecule or expression  
CC product (or fragment) complexed with a human leukocyte antigen (HLA)  
CC molecule and determining the interaction between the agent and the  
CC nucleic acid molecule or the expression product as a determination of the  
CC disorder.  
XX  
SQ Sequence 4660 BP; 1447 A; 1152 C; 995 G; 1066 T; 0 other;

Query Match 5.9%; Score 364.6; DB 21; Length 4660;  
Best Local Similarity 60.2%; Pred. NO. 2.6e-90;  
Matches 689; Conservative 0; Mismatches 429; Indels 27; Gaps 4;  
Qy 155 TGTCAAGATGATGTGGATGGAGAAAATACCTCTGACCCCTCCAGGCAAGACAGAA 214

Db 187 ttctcaacatgagtgagctcgggagcagttcatccgaccctcgccaacccagaccacata 246  
Qy 215 AGCGCAAGGAATGTCTGACCACTGGACCCAGCCCCAAAAGAACATGAAAACGTA 274  
Db 247 agaggaagaagtcgccc-----atgtgacacatctgcaatcaagcaggaagaaagggc 297  
Qy 275 ATCGTGAAACAGGAATAATATATAGAGAACTTGCAAGATTGTGATTTTTCGAATTTTA 334  
Db 298 gcaggagagcaagaataaataatattagagaacatagctgagttacgtctgcgcaacata 357  
Qy 335 ATGATATAGCAACTTTAACTTCAAACTGCAAAATGNGCAATCTTAAATAAAGCTGGA 394  
Db 358 gtgacatlgacagcttgagtgatgtaaaccaagcaaatgcaagatttgaagaagaacagtcg 417  
Qy 395 AGCAAAATTCGTACATCAAAAGAACAGAGAAAGAGCAGCTGCCACATGATGAGTGC 454  
Db 418 atcagatcacgtatgagagaatgagacaagaagaatacaacatgcatgacgtac 477  
Qy 455 AGAAGTCAGATGTATCCTCTACAGGCGAGGTGTCATCGACAAGATCCGTGGGCGCTA 514  
Db 478 agaatacagacatctcatcaagtagtcaagagtagatagaagaagaatccttgggacccc 537  
Qy 515 TGATGCTTGAGGCCCTTGATGGTGTCTTCTTGTAGTGAACCTGGAAGGCAACGTTGTG 574  
Db 538 ttctttggaggtcttgatgagatttcttctgtgtgaaactgtgaaggagaaattgtat 597  
Qy 575 TTGTGTGAGAAATGTGACACAGTATCTAAGGTATTAACCAAGAGAGCTGATGAACAAA 634  
Db 598 ttgtctcagagaatctgaaccagctacttaggttacaatcagggagaaattaatgatacca 657  
Qy 635 GTGATATAGCATCTTGATGTTGGGACACACAGGAATTTGCAAAAACCTGTGCAAA 694  
Db 658 gcgtctcaagacatctcgaagtcggggatcatcgagaatttgaagaatcgtctaccaa 717  
Qy 695 AGTCTATAGTAAATGGGGATCTTGCTGCGCAACCTCCGAGCGGGAACAGCATACT 754  
Db 718 aatacactagtaaatgaggttctcttgctcaaggcgaacacgcgaataatgcataact 777  
Qy 755 TCAATTTGCGAGTCCTGGTAAACCTTTACCTGATTCAGAAAGAGAGGCTCATATACC 814  
Db 778 ttaactgcagatgtcattatccacc-----tccagtagacggcgccagagacc 828  
Qy 815 AGGAAGCTCATCAGAAATATGAACATATGACGTCCTTCGCTCTCACAACCAATGCCA 874  
Db 829 aagaagcttgcacagcgtatgaaatgaaatgagtgcttcaactgtgtcagcagaataatcaa 888  
Qy 875 TCAAGAAGAGAGAGAAATTTGCAATTCGATTCGATTTGGCGCAAGAGATTGCCA 934  
Db 889 ttcaagagatgagagaagatttccagtcacatgtctgatttgcacagcgcatctc 948  
Qy 935 TGAAGAAAGACAGTCTTCCCTCATCAGAAATTTTACCTACCTCCAGAGATCTCCAG 994  
Db 949 g-----cctcagcatlatacgggtgtgagaaatccttataagcaagaagatactacag 1002  
Qy 995 GCAAGATCACCTCTCTGATATCCAGCACATGAAAGACAGCCATTAACACAGCTGGGAGG 1054  
Db 1003 gtaaatcatctcatgtactagttccctgagagctgctggaagaaactgttgggaag 1062  
Qy 1055 ACCGTGTAAGAGAGTGTTCAGAAATTCATGCGCAGCAGAGAAAGAACTGTGTCCT 1114  
Db 1063 attagtagagaaatgcatlatacttcttcccaactcagggagaaac---acctt 1119  
Qy 1115 ATGCTAAGAGCATCATCATGATGATGATGAGACAGATTGGCATTCAGTCAATCTATC 1174  
Db 1120 atgcacagacagctgttccaagaagtgatgacgtgtgacacgtccctccctcata 1179  
Qy 1175 GTTTTCTCTGTGATGAGCGCATCTTGTGCTGCAACAACGAAGCAAAATCTACCGTT 1234  
Db 1180 gatcatatgaaatgagagcaaatgcttagcgccacacaaatgtaaaacttgcacc 1239  
Qy 1235 CTCAGACTACTAATGAACCTCACTTGAATATCTTTACATATCTCTTCACAGAGACAGA 1294



Db 1151 99 1152

RESULT 9  
AAK84490  
ID AAK84490 standard; DNA; 3806 BP.  
XX  
AC AAK84490;  
XX  
DT 07-NOV-2001 (first entry)  
XX  
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:39302.  
XX  
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
KW cytostatic; gene therapy; vaccine; metastasis; ds.  
XX  
OS Homo sapiens.  
XX  
PN W0200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US01354.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217487.  
PR 14-JUL-2000; 2000US-0218280.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 23-AUG-2000; 2000US-0226686.  
PR 23-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.

PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.

PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250160.  
PR 05-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI; 2001-483426/52.  
DR  
XX  
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
PT useful for preventing, diagnosing and/or treating cancers and  
PT metastasis -  
XX  
XX  
XX Disclosure; SEQ ID NO 39302; 3071pp + Sequence Listing; English.  
XX  
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)  
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic  
CC activity, and can be used in gene therapy and vaccine production. (I)  
CC proteins and polynucleotides may be used in the prevention, diagnosis and  
CC treatment of diseases associated with inappropriate (I) expression. For  
CC example, they may be used to treat disorders associated with decreased  
CC expression by rectifying mutations or deletions in a patient's genome  
CC that affect the activity of (I) by expressing inactive proteins or to  
CC supplement the patient's own production of (I). Additionally, (I)  
CC polynucleotides may be used to produce the secreted (I), by inserting  
CC the nucleic acids into a host cell and culturing the cell to express the  
CC protein. (I) proteins and polynucleotides may be used to prevent,  
CC diagnose and treat immune/haematopoietic-related diseases, especially  
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703  
CC to AAK87694 represent human immune/haematopoietic antigen genomic  
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169  
CC represent sequences used in the exemplification of the present invention.  
XX  
SQ Sequence 3806 BP; 1149 A; 672 C; 773 G; 1212 T; 0 other;

Query Match 3.1%; Score 189; DB 22; Length 3806;  
Best Local Similarity 100.0%; Pred. No. 2.4e-41;  
Matches 189; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 702 AGTAATGGGGGATCTTGCTGGCGAAGCTCGAGCGGAGACAGCATACCTTCAATTG 761  
DB 3618 agtaaaagggggatcttggtgctggcgaaacctcgagcggaacacgacatcctcaatg 3677  
QY 762 TCGGATGCTGGTAAACCTTACCTGATTCAGAGAGAGAGGCTCAGATTAACGAGAAC 821  
DB 3678 tcggatgctgtaaaaccttacctgattcagaagaggaagtcacatatacagaagac 3737  
QY 822 TCATCGAAGAAATATGAACCTATGTCAGTGTCTGCTCAACCAAAAGTCATCAAGA 881  
DB 3738 tcattcgaaatattgaaactatgcatgctgtcgtctcacaacaaagtcacataaga 3797  
QY 882 AGAAGGAGA 890  
DB 3798 agaagaga 3806

RESULT 10  
AAK84489  
ID AAK84489 standard; DNA: 3815 BP.  
XX

AC AAK84489;  
XX  
XX 07-NOV-2001 (first entry)  
DT  
XX  
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:39301.  
XX  
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
KW cytostatic; gene therapy; vaccine; metastasis; ds.  
XX  
XX Homo sapiens.  
XX  
PN WO200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US01354.  
XX  
XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226881.  
PR 22-AUG-2000; 2000US-0226882.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227824.  
PR 30-AUG-2000; 2000US-0227824.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.



PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249285.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251968.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.

PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI, 2001-483426/52.  
DR  
XX  
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
PT useful for preventing, diagnosing and/or treating cancers and  
PT metastasis -  
XX  
XX Disclosure; SEQ ID NO 39301; 3071bp + Sequence Listing; English.  
XX  
XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)  
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic  
CC activity, and can be used in gene therapy and vaccine production. (I)  
CC proteins and polynucleotides may be used in the prevention, diagnosis and  
CC treatment of diseases associated with inappropriate (I) expression. For  
CC example, they may be used to treat disorders associated with decreased  
CC expression by rectifying mutations or deletions in a patient's genome  
CC that affect the activity of (I) by expressing inactive proteins or to  
CC supplement the patients own production of (I). Additionally, (I)  
CC polynucleotides may be used to produce the secreted (I), by inserting  
CC the nucleic acids into a host cell and culturing the cell to express the  
CC protein. (I) proteins and polynucleotides may be used to prevent,  
CC diagnose and treat immune/hematopoietic-related diseases, especially  
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703  
CC to AAK87694 represent human immune/hematopoietic antigen genomic  
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169  
CC represent sequences used in the exemplification of the present invention.  
XX  
SQ Sequence 3815 BP; 1141 A; 654 C; 785 G; 1235 T; 0 other;

Query Match 3.0%; Score 187.4; DB 22; Length 3815;  
Best Local Similarity 99.5%; Pred. No. 6.6e-41;  
Matches 188; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 702 AGTAAATGGGGATCTGTGCTGGCAACTCCGAGCGGAGACACCATACCTTCAATG 761  
|||||  
DB 3627 agtaaatggggatctgtgctggcaacctccgagcggaacagcaccactcaatg 3686  
OY 762 TCGATGCTGTAAACCTTTACCTGATTCAGAGAGAGAGGTGCTGATTAACACAGAACG 821  
|||||  
DB 3687 tcgatgctgtlaaaaccttaccgtgaltcagaagagaggtctgtlaaccaggaagc 3746  
OY 822 TCATCAGAAATATGAACTATGACAGTCTGCTCTCTCAACCAAGTCCATCAAGA 881  
|||||  
DB 3747 tcattcagaataatgaaactatgcagtgcttgcgtcttcaaccaaagtcattcaaga 3806  
OY 882 AGAAGGAGA 890  
|||||  
DB 3807 agaagagaga 3815

RESULT 11  
AAH04011  
ID AAH04011 standard; cDNA: 767 BP.

XX AAH04011;

XX 26-JUN-2001 (first entry)

XX Human cDNA clone (5'-primer) SEQ ID NO:846.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.



Oy		958	TCTATCGAAGAAATTTTACTACTCGCCAGGATPTCCAAAGGCAGAATCAGCTCTGTGGATACC	1017
Dd		376	tcaaatctgtagtataatgcctgtctgtacttaataagcaagaaccaagtcccttgataacc	435
Oy		1018	AGCACCATGAGAGACGACCATGAACCAAGCTGGGAGACCGCTGGTAAGAGGTGTATTTCAG	1077
Dd		436	agcacatgatgagacgacctatgaaccaggctcgggaggaacctggttaaaaaagttgattccaq	495
Oy		1078	AAGTTTCATGGCGACGACATGAAGAGAAATCTGTGTCCTATGCTTAAGAGGCATCATCATGAA	1137
Dd		496	aagtcacatgcgcagcatgaaggaaacctgtgtcctatgctaagagagcatcatcatgaag	555
Oy		1138	G 1138	
Dd		556	g 556	
<hr/>				
RESULT 13				
ID	AAH99288			
XX	AAH99288	standard; cDNA;	435 BP.	
AC	AAH99288;			
XX				
DE	16-OCT-2001	(first entry)		
DT				
XX				
XX	Human protein encoding	cDNA sequence SEQ ID NO:123.		
KW	Human:	cancer; HIV infection; human immunodeficiency virus;		
KM	antihistaminic;	antiheumatic; antiarthritic; immunosuppressive;		
KM	antibacterial;	endocrine; cardiant; central nervous system; virucide;		
KW	anti-HIV;	fungicide; antimutagen; cardiovascular; antianaemic; anaemia;		
KW	antiallergent;	haemostatic; vulnery;	antitumor; osteopathic; eczema;	
KW	dermatological;	antiallergic; antiaslomatic; antidiabetic; cytostatic;		
KW	neuroprotective;	antidepressant; nootropic; antiparkinsonian; infection		
KM	immunostimulant;	gene therapy; antisense therapy; vaccine; inflammation		
KM	antidiaphnytic;	Rheumatoid arthritis; septic shock; pancreatitis;		
KW	cardiac dysfunction;	neuropathology; cardiac anaphylaxis; autoimmunity;		
KW	genetic disease;	haematopoietic disorder; platelet disorder; asthma;		
KW	thrombocytopoenia;	osteoporosis; severe combined immunodeficiency;		
KW	allergic rhinitis;	diabetes; multiple sclerosis; depression;		
KW	Alzheimer's disease;	Parkinson's disease; neurodegenerative disorder;		
KW	neurological disorder;	ss.		
OS				
XX	Homo sapiens.			
PN	WO200153455-A2.			
XX				
PD	26-JUL-2001.			
XX				
PF	22-DEC-2000; 2000MO-US35017.			
XX				
PR	23-DEC-1999; 99US-0471275.			
PR	21-JAN-2000; 2000US-0488725.			
PR	25-APR-2000; 2000US-0552317.			
XX				
PA	(HYSE-) HYSEQ INC.			
PI				
DR	Tang YT, Liu C, Drmanac RT;			
XX				
DR	WI: 2001-457603/49.			
XX	P-PSDB; AAM25347.			
PT				
CC	Isolated human polynucleotides encoding polypeptides, useful for the			
CC	treatment and diagnosis of e.g. cancer, ulcers and HIV infection -			
PS	Claim 1; Page 354; 1217pp; English.			
XX				
XX	AAH99166 to AAH99904 encode the human proteins given in AAM25225 to			
CC	AAM25963. The proteins can have activities based on the tissues and			
CC	cells they are expressed in, such as: antihistaminic; antiheumatic;			
CC	antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant;			
CC	central nervous system; virucide; anti-HIV; fungicide; antimutagen;			
CC	cardiovascular; antianaemic; antiallergent; haemostatic; vulnery;			
CC				

CC	antitumor; osteopathic; dermatological; antiallergic; antiasthmatic;
CC	antidiabetic; cytosolic; neuroprotective; antidepressant; motropic;
CC	antiParkinsonian; and immunostimulant. The proteins and polynucleotides
CC	encoding them can be used in gene therapy, antisense therapy and vaccine
CC	production. The proteins and polynucleotides are useful for screening for
CC	agonists or antagonists of a protein and for the treatment and diagnosis
CC	of disorders associated with the activity of a protein e.g. inflammation,
CC	rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
CC	neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
CC	infections, autoimmunity, genetic diseases, haematopoietic disorders,
CC	anemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
CC	osteoporosis, severe combined immunodeficiency, eczema, allergic
CC	rinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
CC	Alzheimer's disease, Parkinson's disease, neurodegenerative and
CC	neurological disorders.
SQ	Sequence 435 BP; 126 A; 70 C; 92 G; 147 T; 0 other:
	Query Match                      1.7%;   Score 106;   DB 22;   Length 435;
	Best Local Similarity         100.0%;   Pred. No. 7.9e-19;
	Matches   106;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0
OY	421 GAGAAACGACAGCTGCCACATAGATGAAGTGCAGACAATCAGATGTATCCTTCACAGG 480
Db	1 ggaagaagcagcagctgcacacatagatgaagtgcagaagtcagatgtacctctacaagg 60
OY	481 CAGGGTGTCATGACAGAAGTAGCGTGGGGCCTTAGATATCCTTGAGG 526
Db	61 cagggtgtcatcgacaagaatgcgctggggccttagatattccttgagg 106
RESULT 14	
AAx80993	
ID	AAx80993 standard; DNA: 3361 BP.
XX	
AC	AAx80993;
XX	
DT	06-SEP-1999 (first entry)
XX	
DE	Mouse steroid receptor coactivator-3 (SRC-3) partial DNA sequence.
XX	
KW	Steroid receptor coactivator-3; SRC-3; transcription enhancement; human;
KM	hormone response element; estrogen receptor alpha; neoplastic state;
KW	mouse; ss.
XX	
OS	Mus sp.
XX	
PN	WO9932621-A2.
XX	
PD	01-JUL-1999.
XX	
PX	01-DEC-1998; 98WO-US25478.
PX	
PR	22-DEC-1997; 97US-0068511.
XX	
PA	(AMHP ) AMERICAN HOME PROD CORP.
XX	
PI	Frali DE, Lyttle CR, Suen C;
XX	
DR	WPI: 1999-405171/34.
XX	
PT	Nucleic acid encoding a steroid receptor co-activator-3, useful for
PP	determining the neoplastic states of cells in humans or animals
XX	
P5	Example 9; Page 73-75; 75pp; English.
XX	
CC	The invention relates to a human steroid receptor coactivator-3 (SRC-3).
CC	Host cells transformed with vectors comprising the SRC-3 gene can be used
CC	for the recombinant production of the SRC-3 protein. SRC-3 can be used to
CC	enhance the transcription of genes, which are regulated by hormone
CC	response elements, e.g. estrous receptor alpha. SRC-3 can be used in
CC	methods to screen for compounds that bind to SRC-3 or which modulate

CC formation of SRC-3 transcription complexes. Determining the amount of  
CC SRC-3 in a sample is also useful in a method to detect overexpression of  
CC SRC-3 and therefore determining the neoplastic states of cells in humans  
CC or animals. The present sequence represents a partial DNA sequence of  
CC mouse SRC-3.  
XX  
SQ Sequence 3361 BP; 872 A; 684 C; 780 G; 1025 T; 0 other:

Query Match 1.5%; Score 95.4; DB 20; Length 3361;  
Best Local Similarity 72.8%; Pred. No. 2.9e-15;  
Matches 123; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

OY 520 CTTGAGCCCTTGAGGGGTTCTTTGTAGTGACCTGGAGGCAACGTTGTTGTG 579  
|| |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | ||||  
DB 2018 cctgaagcaactgattggttccgttctgtgtgacatcgatgaaacattgtatccgtg 2077  
OY 580 TCAGAGAAAGTGCACACATGATCTAAGGTATACCAAGAGAGCTGATGAACAAAGTGA 639  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 2078 tcagaaatgctcacacagatctgcacacagcagagagacctggttacaacagtgctc 2137  
OY 640 TATGACATCTGCATGTGGGACACACGAGATTGTCTCAAAAACCTGC 688  
|| ||||| |||| | |||| | |||| | |||| | |||| | |||| | |||| | ||||  
DB 2138 tacagattcttacatgacagacgcggaagatttcttaaacacttac 2186

RESULT 15  
ABA50553/c  
ID ABA50553 standard; DNA; 330 BP.  
XX ABA50553;

XX 01-FEB-2002 (first entry)  
DE Human breast cell single exon nucleic acid probe #9248.  
XX Human breast; single exon probe; gene expression; breast;  
KM disease; cancer; ss.  
XX Homo sapiens.  
OS  
XX WO200157271-A2.  
PN  
XX 09-AUG-2001.  
PD  
XX 30-JAN-2001; 2001WO-US00662.  
PF  
XX 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.  
PA  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
PI  
XX WPI; 2001-496933/54.  
DR  
XX  
XX

PT New spatially-addressable set of single exon nucleic acid probes,  
PT useful for measuring gene expression in sample derived from human  
PT breast, comprises number of single exon nucleic acid probes -  
XX  
XX  
PS Claim 4; SEQ ID NO 9248; 327bp + sequence listing; English.

XX The invention relates to a spatially-addressable set of single exon  
CC nucleic acid probes for measuring gene expression in a sample derived  
CC from human breast and BT 474 cells. The method involves contacting  
CC the probes with a collection of detectably labeled nucleic acids  
CC derived from mRNA of human breast, and then measuring the label  
CC bound to each probe of the microarray. The probes are useful for

CC verifying the expression of regions of genomic DNA predicted to  
CC encode proteins. They are useful for gene discovery, and for  
CC determining predisposition and/or prognosing breast disease. Gene  
CC expression analysis is useful for assessing the toxicity of chemical  
CC agents on cells. The microarray of this invention presents a far greater  
CC diversity of probes for measuring gene expression, with far less bias  
CC than expressed sequence tag microarrays. The method is suitable for  
CC rapid production of functional information from genomic sequence. The  
CC present sequence is a single exon nucleic acid probe of the invention.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 330 BP; 81 A; 80 C; 73 G; 96 T; 0 other;

Query Match 1.1%; Score 70.6; DB 22; Length 330;  
Best Local Similarity 62.2%; Pred. No. 5e-09; Matches 130; Conservative 0; Mismatches 79; Indels 12; Gaps 2;

OY 2193 TACACATGGAACCTCGCTCAAGAGAGCATAAATTTTGCACAGACTCTTCAGAGACAG 2252  
|| |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | ||||  
DB 242 TATGATGCGGTGCTACTGTTCAAGAGAGACACCGGATTTTGCACAGACTCTTCAGAGATG 183  
OY 2253 CAGTCCCTGCTGACTGGCCCAAGTTAACAGCAGACACGCAAGACCTGAGCCA 2312  
|| |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | ||||  
DB 182 GAATTCACAGCTGAGTGAAGATTAAGTACGAGAGACCTGAGAGACACAGCAG--- 124  
OY 2313 GAGTCCAGCAGCAGACGCTCCGATCAGAGTACTATTAAACAGAGCGGTGAGCCC 2372  
|| |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | ||||  
DB 125 -----CAGTATTAATCTTTTGTGGAGCGAATAATTTGTCACAGAGACAGCAGTAACTGC 72  
OY 2373 CAAGAGAAAGAG--ATGCACTACTTGCCTATTGCTAGATMAAGATGATCTAAAGA 2429  
||||| |||| | ||||| |||| | |||| | |||| | |||| | |||| | |||| | ||||  
DB 71 TAAGAGAGAGAGATAATAGCACTTCTTAGATATCTGCTGAGACAGAGATGATCTAGTGA 12  
OY 2430 T 2430  
DB 11 T 11

RESULT 16  
ABA68512/c  
ID ABA68512 standard; DNA; 330 BP.  
XX ABA68512;

XX 01-FEB-2002 (first entry)  
DE Human foetal liver single exon nucleic acid probe #16817.  
XX Human foetal liver; single exon probe; gene expression; single exon nucleic acid probe; ss.  
KM Human; foetal liver; gene expression; single exon nucleic acid probe; ss.  
XX Homo sapiens.  
OS  
XX WO200157277-A2.  
PN  
XX 09-AUG-2001.  
PD  
XX 30-JAN-2001; 2001WO-US00669.  
PF  
XX 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.  
PA  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
PI  
XX

XX



XX 30-JAN-2001; 2001WO-US00667.  
 XX  
 PF 04-FEB-2000; 2000US-0180312.  
 XX 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 XX 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 04-OCT-2000; 2000US-0236359.  
 XX 27-SEP-2000; 2000US-0234687.  
 PR 04-OCT-2000; 2000US-0236359.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX Penn SG, Hanzel DK, Chen W, Rank DR.  
 XX WPI; 2001-483446/52.  
 DR  
 XX  
 PT Single exon nucleic acid probes for analyzing gene expression in human  
 XX brains -  
 PS Example 4; SEQ ID NO: 16870; 650bp + Sequence Listing; English.  
 XX  
 CC The present invention provides a number of single exon nucleic acid  
 CC probes which are derived from genomic sequences expressed in the human  
 CC brain. They can be used to measure gene expression in brain cell samples,  
 CC which may enable the diagnosis and improved treatment of nervous system  
 CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
 CC epilepsy and cancers. The present sequence is one of the probes of the  
 CC invention.  
 XX  
 SQ Sequence 330 BP; 81 A; 80 C; 73 G; 96 T; 0 other;

Query Match 1.1%; Score 70.6; DB 22; Length 330;  
 Best Local Similarity 62.2%; Pred. No. 5e-09;  
 Matches 150; Conservative 0; Mismatches 79; Indels 12; Gaps 2;  
 QY 2193 TACACATGGAACCTCGCTCAAGAGAGCATAAATTTCACAGACTCTTGACAGACAG 2252  
 DB 242 TATGATGATGGTCACTGTTACAGAGAGCACCAGATTTCACAAAGTTGCTGAGAAATGG 183  
 QY 2253 CAGTCCCTCGTGGAGCTTGCCCAAGTTAACAGAGAGCCACAGCCAAAGACCTGAGCCA 2312  
 DB 182 GAATTCACCGAGTGAAGTACCAAGATTACTGCAAGAGCCACAGGAAAGACACCCAG-- 124  
 QY 2313 GGAGTCCAGCAGCAGCAGCTCTGATCAGAGTACTATTAAACAGAGCCGCGTGAAGCC 2372  
 DB 125 -----CAGTATTAACCTTCTGTGGGAGCGAATGTTGTCAAGCAGAGCAGCTAAGTCC 72  
 QY 2373 CAGAGAGAAAGAG--AATGCACTACTTCGCTATTGCTAGATTAAGATGATCTAAAGA 2429  
 DB 71 TAAGAGAGAGAGAGATTAATGCACTCTTAGATACCTGCTGACAGAGGATGATCTTAGTGA 12  
 QY 2430 T 2430  
 DB 11 T 11

RESULT 19  
 ID AAK42647 standard; DNA: 330 BP.  
 XX AAK42647;  
 AC  
 XX  
 DT 06-NOV-2001 (first entry)  
 XX  
 DE Human bone marrow expressed single exon probe SEQ ID NO: 17204.  
 XX  
 KW Human; bone marrow expressed exon; gene expression analysis; probe;  
 XX microarray; cancer; leukaemia; lymphoma; myeloma; ss.  
 OS Homo sapiens.  
 XX

PN W0200157276-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 30-JAN-2001; 2001WO-US00668.  
 XX  
 XX 04-FEB-2000; 2000US-0180312.  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 04-OCT-2000; 2000US-0236359.  
 XX 27-SEP-2000; 2000US-0234687.  
 PR 04-OCT-2000; 2000US-0236359.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX Penn SG, Hanzel DK, Chen W, Rank DR.  
 XX WPI; 2001-488900/53.  
 DR  
 XX  
 PT Human genome-derived single exon nucleic acid probes useful for  
 XX analyzing gene expression in human bone marrow -  
 PS Example 4; SEQ ID NO: 17204; 658bp + Sequence Listing; English.  
 XX  
 CC The present invention provides a number of single exon nucleic acid  
 CC probes which are derived from genomic sequences expressed in the human  
 CC bone marrow. They can be used to measure gene expression in bone marrow  
 CC samples, which may enable the improved diagnosis and treatment of cancers  
 CC such as lymphoma, leukaemia and myeloma. The present sequence is one of  
 CC the probes of the invention.  
 XX  
 SQ Sequence 330 BP; 81 A; 80 C; 73 G; 96 T; 0 other;

Query Match 1.1%; Score 70.6; DB 22; Length 330;  
 Best Local Similarity 62.2%; Pred. No. 5e-09;  
 Matches 150; Conservative 0; Mismatches 79; Indels 12; Gaps 2;  
 QY 2193 TACACATGGAACCTCGCTCAAGAGAGCATAAATTTCACAGACTCTTGACAGACAG 2252  
 DB 242 TATGATGATGGTCACTGTTACAGAGAGCACCAGATTTCACAAAGTTGCTGAGAAATGG 183  
 QY 2253 CAGTCCCTCGTGGAGCTTGCCCAAGTTAACAGAGAGCCACAGCCAAAGACCTGAGCCA 2312  
 DB 182 GAATTCACCGAGTGAAGTACCAAGATTACTGCAAGAGCCACAGGAAAGACACCCAG-- 124  
 QY 2313 GGAGTCCAGCAGCAGCAGCTCTGATCAGAGTACTATTAAACAGAGCCGCGTGAAGCC 2372  
 DB 125 -----CAGTATTAACCTTCTGTGGGAGCGAATGTTGTCAAGCAGAGCAGCTAAGTCC 72  
 QY 2373 CAGAGAGAAAGAG--AATGCACTACTTCGCTATTGCTAGATTAAGATGATCTAAAGA 2429  
 DB 71 TAAGAGAGAGAGAGATTAATGCACTCTTAGATACCTGCTGACAGAGGATGATCTTAGTGA 12  
 QY 2430 T 2430  
 DB 11 T 11

RESULT 20  
 ID AAI23403 standard; DNA: 330 BP.  
 XX AAI23403;  
 AC  
 XX  
 DT 12-OCT-2001 (first entry)  
 XX  
 DE Probe #13336 for gene expression analysis in human cervical cell sample.  
 XX  
 KW Probe; human; microarray; gene expression; cervical epithelial cell;  
 XX cervical cancer; ss.  
 XX

OS Homo sapiens.  
XX  
PN WO200157278-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US00670.  
XX  
PR 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
DR WPI; 2001-488901/53.  
XX  
PT Human genome-derived single exon nucleic acid probes useful for  
XX analyzing gene expression in human cervical epithelial cells -  
XX  
XX  
XX Claim 25; SEQ ID No 13336; 487bp; English.  
XX  
XX The present invention relates to human single exon nucleic acid probes  
CC (SENPs). The present sequence is one such probe. The SENPs are derived  
CC from human HeLa cells. The SENPs can be used to produce a single exon  
CC microarray, which can be used for measuring human gene expression in a  
CC sample derived from human cervical epithelial cells. By measuring gene  
CC expression, the probes are therefore useful in grading and/or staging  
CC of diseases of the cervix, notably cervical cancer.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at [ftp://wipo.int/pub/published\\_pcl\\_sequences](http://wipo.int/pub/published_pcl_sequences).  
XX  
SQ Sequence 330 BP; 81 A; 80 C; 73 G; 96 T; 0 other;

Query Match 1.1%; Score 70.6; DB 22; Length 330;  
Best Local Similarity 62.2%; Pred. No. 5e-09;  
Matches 150; Conservative 0; Mismatches 79; Indels 12; Gaps 2;

QY 2193 TACACATGGAACCTCGCTCAAGAGACATTAATAATTTGACACAGACTCTTGACAGACAG 2252  
DB 242 TATGCATGGGTGCTACTGTACAAAGAACACCGGATTTTGCAAAAGTTGCTGCAGAAATGG 183  
QY 2253 CAGTTCCCTGTGAGCTTGGCCAAAGTTAAACAGAGAACCCACAGCAAGCACTGAGCCA 2312  
DB 182 GAATTCACCAAGCTGAGGAGCAAGATTACTGCAGAACCCACTGGGAAAGACACCAG--- 124  
QY 2313 GGAGTCACGACGACAGCTCTCGATCAGAGTGAAGTAAACAGAGCCGGTGAGCC 2372  
DB 125 -----CAGTATTAACCTTTGTTGGGAGCAAAATGTTGTAAGCAGAGAGCTAAGTCC 72  
QY 2373 CAAGAAGAAAGAG---AATGCACTACTTCTGCTATTTGCTAATTAAGTGTACTAAGA 2429  
DB 71 TAAAGAAAGAGAGAAATATGCACTTTAGATTAACCTGCTGAGAGGATGATCCTAAGTGA 12  
QY 2430 T 2430  
DB 11 T 11

RESULT 21  
AAI48724/c  
ID AAI48724 standard; DNA: 330 BP.  
XX  
AC AAI48724;  
XX  
DT 17-OCT-2001 (first entry)

XX  
DE Probe #17410 used to measure gene expression in human placenta sample.  
XX  
KW Probe; microarray; human; placenta; antenatal diagnosis;  
KW genetic disorder; ss.  
XX  
XX Homo sapiens.  
XX  
PN WO200157272-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US00663.  
XX  
PR 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
DR WPI; 2001-488897/53.  
XX  
PT Human genome-derived single exon nucleic acid probes useful for  
XX analyzing gene expression in human placenta -  
XX  
XX  
XX Claim 25; SEQ ID No 17410; 654bp; English.  
XX  
XX The present invention relates to single exon nucleic acid probes (SENPs).  
CC The present sequence is one such probe. The probes are useful for  
CC producing a microarray for predicting, measuring and displaying gene  
CC expression in samples derived from human placenta. The probes are useful  
CC for antenatal diagnosis of human genetic disorders.  
XX  
SQ Sequence 330 BP; 81 A; 80 C; 73 G; 96 T; 0 other;

Query Match 1.1%; Score 70.6; DB 22; Length 330;  
Best Local Similarity 62.2%; Pred. No. 5e-09;  
Matches 150; Conservative 0; Mismatches 79; Indels 12; Gaps 2;

QY 2193 TACACATGGAACCTCGCTCAAGAGACATTAATAATTTGACACAGACTCTTGACAGACAG 2252  
DB 242 TATGCATGGGTGCTACTGTACAAAGAACACCGGATTTTGCAAAAGTTGCTGCAGAAATGG 183  
QY 2253 CAGTTCCCTGTGAGCTTGGCCAAAGTTAAACAGAGAACCCACAGCAAGCACTGAGCCA 2312  
DB 182 GAATTCACCAAGCTGAGGAGCAAGATTACTGCAGAACCCACTGGGAAAGACACCAG--- 124  
QY 2313 GGAGTCACGACGACAGCTCTCGATCAGAGTGAAGTAAACAGAGCCGGTGAGCC 2372  
DB 125 -----CAGTATTAACCTTTGTTGGGAGCAAAATGTTGTAAGCAGAGAGCTAAGTCC 72  
QY 2373 CAAGAAGAAAGAG---AATGCACTACTTCTGCTATTTGCTAATTAAGTGTACTAAGA 2429  
DB 71 TAAAGAAAGAGAGAAATATGCACTTTAGATTAACCTGCTGAGAGGATGATCCTAAGTGA 12  
QY 2430 T 2430  
DB 11 T 11

RESULT 22  
AAI09030/c  
ID AAI09030 standard; DNA: 330 BP.  
XX  
AC AAI09030;  
XX

DT 09-OCT-2001 (first entry)  
XX  
DE Probe #9021 used to measure gene expression in human breast sample.  
XX  
XX Probe: human; breast disease; breast cancer; development disorder; ss;  
KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.  
XX  
OS Homo sapiens.  
XX  
PN WO200157270-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 29-JAN-2001; 2001WO-US00661.  
XX  
PR 04-FEB-2000; 2000US-0180312.  
XX  
PR 26-MAY-2000; 2000US-0207456.  
XX  
PR 30-JUN-2000; 2000US-0608408.  
XX  
PR 03-AUG-2000; 2000US-0632366.  
XX  
PR 21-SEP-2000; 2000US-0234687.  
XX  
PR 27-SEP-2000; 2000US-0236359.  
XX  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
DR WPI; 2001-476286/51.  
XX  
XX  
PT Novel single exon nucleic acid probe used to measuring gene expression  
XX in a human breast -  
XX  
PS Claim 25; SEQ ID NO 9021; 322pp; English.  
XX  
CC The present invention relates to novel single exon nucleic acid probes.  
CC The present sequence is one such probe. The probes are useful for  
CC measuring human gene expression in a human breast sample, where the probe  
CC hybridises at high stringency to a nucleic acid expressed in the human  
CC breast. The probes are useful for predicting, diagnosing, grading,  
CC staging, monitoring and prognosing diseases of the human breast,  
CC particularly those diseases with polygenic aetiology. The diseases  
CC include: breast cancer, disorders of development, inflammatory diseases  
CC of the breast, fibrocystic changes, proliferative breast disease and  
CC non-carcinoma tumours.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 330 BP; 81 A; 80 C; 73 G; 96 T; 0 other;  
SQ

Query Match 1.1%; Score 70.6; DB 22; Length 330;  
Best Local Similarity 62.2%; Pred. No. 5e-09;  
Matches 150; Conservative 0; Mismatches 79; Indels 12; Gaps 2;

QY 2193 TACACATGGAACCTGCTCAAGAGAGCATATAATTTTGACAGACTCTTGACGAGACAG 2252  
II IIII  
DB 242 TATGCTGCGGTGCTTTACAGAGAGCACCGGATTTTGACACAAAGTCTGCGAAGATG 183  
II IIII  
QY 2253 CAGTTCGCCCTGTGAGCTTGCCCAAGTTACAGAGAGACGACGGAAGAGCTGAGCCA 2312  
II IIII  
DB 182 GAATTCACCAAGCTGAGGTGCTGATCAGAGATTAAGTCAAGAGCCAGAGACCCAG--- 124  
II IIII  
QY 2313 GGAGTCCACAGACAGCTGCTGATCAGAGATTAAGTCAAGAGCCGCTGAGCCC 2372  
II IIII  
DB 125 -----CATTAATACTTCTTGTGGGACGGAATGTTGTCMAAGCAGGACGAGTAAGTCC 72  
II IIII  
QY 2373 CAAGAAGAAAGAG---AATGACACTACTTGCTATTTGCTAGATTAAGATGATTAAGATA 2429  
II IIII  
DB 71 TAAGAAGAGAGAAATTAATGACACTTCTTAGATACCTGCTGAGACAGGATGATGATCCTAGTGA 12  
II IIII  
QY 2430 T 2430  
I

DB 11 T 11  
RESULT 23  
ID AAT84543  
ID AAT84543 standard; cDNA; 3388 BP.  
XX  
XX AAT84543;  
XX  
AC 04-DEC-1997 (first entry)  
XX  
DT  
XX  
DE Human steroid receptor coactivator-1 (SRC-1) cDNA.  
XX  
XX Steroid receptor coactivator-1; SRC-1; molecular switch;  
KW gene therapy; transgenic animal; ss.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT CDS 57..3242  
FT /\*tag= a  
XX  
PN WO9710337-A1.  
XX  
PD 20-MAR-1997.  
XX  
XX  
PF 20-AUG-1996; 96WO-US13482.  
XX  
PR 15-SEP-1995; 95US-0003784.  
XX  
PA (BAYU) BAYLOR COLLEGE MEDICINE.  
XX  
PI O'Malley BO, Onate S, Tsai M, Tsai SY;  
XX  
DR WPI: 1997-202233/18.  
XX  
DR P-PSDB; AAM26370.  
XX  
PT Steroid receptor coactivator-1 polypeptide and DNA - regulated using  
XX a molecular switch, used in gene therapy  
XX  
PS Claim 3; Fig 1; 116pp; English.  
XX  
XX This cDNA sequence encodes human steroid receptor coactivator-1  
CC (SRC-1) (AAM26370), a novel protein required for human progesterone  
CC receptor (hPR) transcriptional function. The 3' end of the cDNA  
CC was obtained by screening a fibroblast library with a partial SRC-1  
CC clone obtained from a two-hybrid screening, and the 5' end from a  
CC HeLa cell library by PCR using nested primers (see AAT84546). SRC-1  
CC nucleic acid may be used to treat a SRC-1 related disease by  
CC introducing the sequence into a host cell and infusing the cells  
CC into the patient causing an increase in the transcription of SRC-1  
CC (claimed). A molecular switch can be used to regulate expression  
CC of a nucleic acid cassette incorporating an SRC-1 coding region for  
CC use in gene therapy. Transcription of a target gene can be  
CC decreased by providing a nucleic acid encoding a dominant-negative  
CC inhibitor of a SRC-1 polypeptide in a cell containing the target  
CC gene (also claimed). A nucleic acid probe for detecting SRC-1, and  
CC a transgenic mammal encoding SRC-1 are also claimed.  
XX  
SQ Sequence 3388 BP; 1045 A; 892 C; 706 G; 745 T; 0 other;

Query Match 1.1%; Score 64.8; DB 18; Length 3388;  
Best Local Similarity 52.1%; Pred. No. 1e-06;  
Matches 250; Conservative 0; Mismatches 212; Indels 18; Gaps 4;

QY 4068 TCCAAACTAGGAAATAGTCACAGCACTGATCCAGGCTTTACTGGGCTACGACTGCCCA 4127  
II IIII  
DB 2612 tccaggagcaggaatggtccccaagtgagcgcaacttgcctcctaagccctgg 2671  
II IIII  
QY 4128 GAGCCCACTTATGTACCCCGAATGGACATACAGAGATCCCATGATGATCAACAGCTCTGA 4187  
II IIII  
DB 2672 gagctccatggtgcgcatgcacatccctcctcagaggttctcgtccagcaactcc 2731  
II IIII I





PT for diagnosis and treatment of diseases associated with abnormal  
cytosine methylation -

PS Claim 1; SEQ ID NO 863; 32pp + Sequence Listing: German.

CC The present invention provides a number of human immune system associated  
CC genes which are modified by the methylation of cytosines. The sequences  
CC can be used in the diagnosis and treatment of immune system disorders,  
CC including eye diseases such as retinopathy, neovascular glaucoma and  
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
CC diseases. The present sequence is a gene of the invention.

XX Sequence 6131 BP; 2130 A; 39 C; 1018 G; 2944 T; 0 other;

Query Match 0.9%; Score 58; DB 24; Length 6131;  
Best Local Similarity 48.0%; Pred. No. 0.00012;  
Matches 166; Conservative 0; Mismatches 180; Indels 0; Gaps 0;

QY 5778 AAAATCCATTATATATGTAACAGCAATATTAGTTGAGAACTTATGTTTAAATT 5837  
DB 5318 aatatttataataataataataataataataataataataataataattttt 5377  
QY 5838 GTATAATTTTGTGAGCTATACATATTGTGAATTGACTCAAAAATGAGTACTTCAGTA 5897  
DB 5378 gaaaattatgatttggaataattatttttttgtaattattataaagtaattatttt 5437  
QY 5898 TTAATTAATATCTTCATAGCAATGTCCTCAAGGCTTTTGTAAAGATATCAATGC 5957  
DB 5438 ttcttggttggaataataataataataataataataataataataataataata 5497  
QY 5958 CTGATTTAGACCTAATTTTGTAGACCTTAAAGCTTTTATTTTAAACCTTGATCTGCG 6017  
DB 5498 atagattataataaagttataataataataataataataataataataataata 5557  
QY 6018 TTATAAGTATTTATCTTAATGATATGATGAGCGCGTGAAGAACCAATCTTGATT 6077  
DB 5558 ttttttaataataataataataataataataataataataataataataataata 5617  
QY 6078 TTTATATGTTTATCTTCTTAAATGAACTTGAAGAACTTCATCAT 6123  
DB 5618 ttctgataagattatttctgattatttagaataataataataataataata 5663

RESULT 26  
AAS46423

ID AAS46423 standard; DNA; 7135 BP.

XX AAS46423;

DT 18-DEC-2001 (first entry)

DE Tumour suppressor gene derived chemically modified sequence #145.

XX Human; tumour suppressor gene; oncogene; antitumour; cytostatic;

KW cancer; tumour; Cpg dinucleotide; single-nucleotide polymorphism; SNP;

KM cytosine methylation; ds.

OS Homo sapiens.

PN WO200168912-A2.

PD 20-SEP-2001.

PF 15-MAR-2001; 2001WO-EP02955.

XX 15-MAR-2000; 2000DE-1013847.

PR 06-APR-2000; 2000DE-1019058.

PR 07-APR-2000; 2000DE-1019173.

PR 30-JUN-2000; 2000DE-1032529.

PR 01-SEP-2000; 2000DE-1043826.

XX (EPiG-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

DR WPI: 2001-602752/68.

XX Fragments of chemically modified genes associated with tumour suppressor  
XX genes and oncogenes, useful in designing primers and probes for  
XX analysing diseases associated with cytosine methylation state e.g.  
XX cancer -

PS Claim 1; SEQ ID NO 145; 27pp; English.

CC The invention relates to a nucleic acid comprising a sequence of 18  
CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with  
CC bisulphite, of genes associated with tumour suppression and  
CC oncogenes having a sequence taken from 536 (actually 533 since  
CC numbers 408, 458 and 500 are missing from the sequence listing) sequences  
CC (Ss) and sequences complementary to (Ss). The nucleic acid may be a  
CC peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may  
CC form part of a set of probes for detecting the cytosine methylation state  
CC and/or single nucleotide polymorphisms and also to be used in an  
CC array for analysing diseases associated with Cpg dinucleotides e.g.  
CC cancers and tumours. The probes can also be used in a method for  
CC ascertaining genetic and/or epigenetic parameters for the diagnosis  
CC and/or therapy of existing diseases or the predisposition to specific  
CC diseases, by analysing cytosine methylations. The parameters may be  
CC compared to another set of genetic and/or epigenetic parameters, the  
CC differences serving as basis for diagnosis and/or prognosis events which  
CC are advantageous to patients. The present sequence is one of the  
CC 533 genomic sequences derived from tumour suppressor genes and  
CC oncogenes.

CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct-sequences.

XX Sequence 7135 BP; 1713 A; 118 C; 1466 G; 3838 T; 0 other;

Query Match 0.9%; Score 56.8; DB 22; Length 7135;  
Best Local Similarity 49.4%; Pred. No. 0.00029;  
Matches 174; Conservative 0; Mismatches 177; Indels 1; Gaps 1;

QY 5805 ATATTTAGTTGAGAACTTATGTTTAAATTTGTAATTTTGTGAGGTATACATATT 5864  
DB 1917 atgttttattatgttt 1976  
QY 5865 GTGGAATTGACTCAAAAATGAGTACTTCAGTATTAATTTAGATATCTTCATAGCAATGT 5924  
DB 1977 gtttagttta-ttgaataataatttttataaggttagtaagttatttttttttttttt 2035  
QY 5925 CTCTTAAGGCTTTTGTAAAGGATATCAATGCTTGAATTTAGACCTAATTTAGACTTA 5984  
DB 2036 tattaagttcgggttttctgtatttgaatttgaatgtagtaattagtttagtttt 2095  
QY 5985 AGACTTTTATTTTCTTAAACCTTGTGATTCGTTTAAATGATTTATTTATTTATTTATG 6044  
DB 2096 att 2155  
QY 6045 ATATGCAAGCGCTGTAGGAACCAATTTGATTTTAAATGTTTAAATTTATTTATTTATG 6104  
DB 2156 ttt 2215  
QY 6105 AACCTTAGAAAGATACATGTTTACTAAGCAAGCACTTTTATGTTGTTT 6156  
DB 2216 tagtttttaatttgaaatgattttagtttagttttagaagtttttttttttttttttt 2267

RESULT 27

ABJ32911  
ID ABLJ32911 standard; DNA; 8946 BP.

XX	ABLJ32911;
AC	
XX	
DT	26-MAR-2002 (first entry)
XX	
DE	Human immune system associated gene SEQ ID NO: 884.
XX	
KM	Human; immune system disease; cytosine methylation; antiasthmatic;
KM	antiartherosclerotic; antihaemic; cytosolic; nootropic;
KM	neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KM	antihaumatic; antiarthritic; antidiabetic; antipsoriatic;
KM	antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KM	acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KM	neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW	gene; ds.
XX	
OS	Homo sapiens.
PN	
PD	WO200200928-A2.
XX	
PD	03-JAN-2002.
XX	
PF	02-JUL-2001; 2001WO-EP07537.
XX	
PR	30-JUN-2000; 2000DE-1032529.
PR	01-SEP-2000; 2000DE-1043826.
XX	
PA	(EPIG-) EPIGENOMICS AG.
XX	
PL	Olek A, Piepenbrock C, Berlin K;
XX	
DR	WPI: 2002-130909/17.
PT	
PT	Nucleic acid comprising fragment of chemically modified gene, useful
PT	for diagnosis and treatment of diseases associated with abnormal
PT	cytosine methylation -
XX	
PS	Claim 1; SEQ ID NO 884; 32pp + Sequence Listing; German.
XX	
CC	The present invention provides a number of human immune system associated
CC	genes which are modified by the methylation of cytosines. The sequences
CC	can be used in the diagnosis and treatment of immune system disorders,
CC	including eye diseases such as retinopathy, neovascular glaucoma and
CC	macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC	leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC	rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC	diseases. The present sequence is a gene of the invention.
XX	
XX	Sequence 8946 BP; 2493 A; 66 C; 1749 G; 4638 T; 0 other;

	Query Match	0.9%	Score 56.4	DB 24	Length 8946
	Best Local Similarity	48.7%	Pred. NO. 0.00044		
	Matches 182	Conservative 0	Mismatches 191	Indels 1	Gaps 1
QY	5719	TGTGTTTTTATATATGTTTATTTTATGACAGTGCACACAGCCGAGTTTGTGAAGTACATA	5778		
Db	5858	tttatctttaaattttttttaaattattttaaataagatataaattttataattgatltta	5917		
QY	5779	AAATCCAAATTATATGTTAAACCAACCAATTAATTAAAGTTGACAGACTTATGTTTTAATGG	5838		
Db	5918	tgtgttaattctcatgcttgctagctagcttctgattctttagtattataaataattataa	5977		
QY	5839	TATATTTTTTGGAGGTGTACATATTTGGGAATTCACACCAAAAT-GAGGTACTTCACGA	5897		
Db	5978	tttagtgcttcttgtaagaatactttcttagatgctaatctgaattataatacttgctgattt	6037		
QY	5898	TTAAATTAGATATCTTCATAGCAATGTCTCTTAAAGGCTTTTGTAAAAGATATCAATGC	5957		
Db	6038	taagaattattttagaaaaaattgggtgagtttaatttaagtttgcttgaagaattttaaga	6097		
QY	5958	CTTGATTGACCTAATTTGTAGACTTAAAGCTTTTATTTTCTAAACCTTGATATCTGC	6017		

Db	6098	gtcagcttcgagcttcttttttgaggaaaaagaattatcttcttgtagatcgctgttaagtcttgc	6157
Qy	6018	TTATAGTCATTTCATCTAATCATATGATATGCAGCGCCTGTAGCAACCAATTCCTTGATT	6077
DB	6158	gttcgagagcttcgaatttgttcttttttgataataattatcttatcttagtgagacttttgattgttt	6217
Qy	6078	TTTTATATGTTTATA 6091           + + + + +	
Db	6218	agttatcttcatla 6231	
RESULT 28			
ID	ABL32322	standard; DNA; 6418 BP.	
XX	ABL32322;		
AC			
XX			
DT	26-MAR-2002	(first entry)	
XX			
DE	Human immune system associated gene SEQ ID NO: 295.		
XX			
KW	Human; immune system disease; cytosine methylation; antiasthmatic;		
KW	antiartherosclerotic; antianaemic; cytosolic; neotropic;		
KW	neuroprotective; anti-HIV; anticonvulsant; ophthalmological;		
KW	antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;		
KW	acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;		
KW	neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;		
KM	gene; ds.		
XX			
OS	Homo sapiens.		
XX			
PN	WO200200928-A2.		
XX			
PD	03-JAN-2002.		
XX			
PF	02-JUL-2001; 2001MO-EP07537.		
XX			
PR	30-JUN-2000; 2000DE-1032529.		
PR	01-SEP-2000; 2000DE-1043826.		
XX			
PA	(EPIG-) EPIGENOMICS AG.		
XX			
PI	Olek A, Piepenbrock C, Berlin K;		
XX			
DR	WPI; 2002-130909/17.		
XX			
PT	Nucleic acid comprising fragment of chemically modified gene, useful		
PT	for diagnosis and treatment of diseases associated with abnormal		
PT	cytosine methylation -		
XX			
PS	Claim 1; SEQ ID NO 295; 32pp + Sequence Listing; German.		
XX			
CC	The present invention provides a number of human immune system associated		
CC	genes which are modified by the methylation of cytosines. The sequences		
CC	can be used in the diagnosis and treatment of immune system disorders,		
CC	including eye diseases such as retinopathy, neovascular glaucoma and		
CC	macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid		
CC	leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,		
CC	rheumatoid arthritis, psoriasis and inflammatory/intestinal bowel		
CC	diseases. The present sequence is a gene of the invention.		
XX			
SO	Sequence 6418 BP; 1842 A; 47 C; 1423 G; 3106 T; 0 other;		
Query Match            0.9%; Score 55.6; DB 24; Length 6418; Best Local Similarity   44.7%; Pred. No. 0.00059; Matches 217; Conservative 0; Mismatches 269; Indels 0; Gaps 0;			
Qy	5670	ACTGCTTTAAATTCATACGATATCACGACAGATTTTAACCTCTGCCATGCTTTTAA 5729	
DB	2364	aatggaaagcttcagtaaataaatatgatgttctttgttggagagcttagctatt 2423	

QY 5730 TTTTGGTTTTAGCAGTGTGCTGACTNAGCCGAAATTTTGTAGGTACATNAAATCCAAATTT 5789  
DB 2424 ttttggtttgaattaaattatgttggttggaagtggttgattagatgattttttttaaaaaa 2483  
QY 5790 AATATGAAACCAACATTAATTTAAGTGTGAGAACTTATGCTTTAATTTGATTAATTTTG 5849  
DB 2484 tttgtttaaaaaatgatttggttttaattagtgatataataatcgatgaggtttt 2543  
QY 5850 TCGAGTATCATATGTGTGCAATTTGACTCAAAAATGAGTACCTTCAATTTAATTAATGATA 5909  
DB 2544 ttgaggaaagttaattatgttgagtaatgtatatttttggatttggatataattttgatt 2603  
QY 5910 TCTTCATACCAATGCTCCTTAAAGTGTTTGTAAAGATATCAATGCTTGAATAGACC 5969  
DB 2604 tataaataatgttttttattgatagtggttttgattgggaaaatttttgtaataatgatt 2663  
QY 5970 TAATTTGTAGACTTAAGACTTTTATTTTCTTAACCTTGATGCTTGCTTAAAGCATT 6029  
DB 2664 tgttttttttattttatttatttataaagtataattatgatttatttatttatt 2723  
QY 6030 TATCTATATCATATGATATGATGACGCCGCTGAGAACCAATTTCTGATTTTATATGTTTA 6089  
DB 2724 tataataatttgatgttgaagatttgggtttatatattttttttttatttatttattt 2783  
QY 6090 TATTTCTTCTTAATGAACCTTAGAAAGACTACATGTTTACTAAGCAGGCCACTTTATGCT 6149  
DB 2784 tattttatgaacgaagaattatttggtttgagtaattgttgagaattatataatttaggt 2843  
QY 6150 TGTTTT 6155  
DB 2844 tttttt 2849

RESULT 29

AAS61073  
ID AAS61073 standard; DNA; 6418 BP.

AAS61073;

29-JAN-2002 (first entry)

Human gene regulation-associated gene oligonucleotide #28.

Human; Gene regulation-associated gene; severe combined immunodeficiency;  
cardiac damage; inflammatory response; Haemophilia; Werner syndrome;  
asthma; HDR syndrome; congenital heart defect; Saethre-Chotzen syndrome;  
renal disease; Preeclampsia; cardiac allograft vascular disease;  
colorectal cancer; thyroid cancer; oesophageal cancer; ds. tumour;  
immunostimulant; cardiac; antiinflammatory; coagulant; antiasthmatic;  
neurotropic; gynecological; anti-tumour; immunosuppressive; cytostatic.

Homo sapiens.

WO200177375-A2.

18-OCT-2001.

06-APR-2001; 2001MO-EP03968.

06-APR-2000; 2000DE-1019058.

07-APR-2000; 2000DE-1019173.

30-JUN-2000; 2000DE-1032529.

01-SEP-2000; 2000DE-1043826.

(EPIG-) EPIGENOMICS AG.

Olek A, Piepenbrock C, Berlin K;

WPI; 2002-017470/02.

New nucleic acid sequences from chemically modified genes associated  
with gene regulation, useful for analysing cytosine methylations for  
diagnosis and therapy of diseases e.g. severe combined immunodeficiency

PT disease  
XX  
PS Claim 1; SEQ ID No 29; 26pp; English.  
XX  
CC The invention relates to 224 nucleic acid sequences comprising at least  
CC 18 bases of a chemically pretreated gene associated with gene regulation  
CC selected from 43 known genes (or complementary sequences). The  
CC chemical pretreatment converts cytosine bases unmethylated at the  
CC 5-position to uracil or another base with hybridisation behaviour  
CC similar to cytosine, to enable analysis of cytosine methylations.  
CC The DNA sequences, oligomers (or sets/array) and method are  
CC useful in the diagnosis of diseases (or predisposition to diseases)  
CC associated with gene regulation and in therapy of such diseases. By  
CC enabling analysis of the cytosine methylation patterns of such genes,  
CC kits are provided. They are especially useful in diagnosis  
CC and therapy of e.g. severe combined immunodeficiency disease, cardiac  
CC disorders, haemophilia, solid tumours and cancer, Werner syndrome,  
CC asthma, HDR syndrome, Saethre-Chotzen syndrome, renal disease,  
CC preeclampsia, graft versus-host disease. The present sequence is a  
CC sequence included in the sequence data for this specification and is  
CC associated with the human gene regulation-associated genes.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 6418 BP; 1842 A; 47 C; 1423 G; 3106 T; 0 other;

Query Match 0.9%; Score 55.6; DB 24; Length 6418;  
Best Local Similarity 44.7%; Pred. No. 0.00059;  
Matches 217; Conservative 0; Mismatches 269; Indels 0; Gaps 0;

QY 5670 AGTGTGTTTAAATGATGATGATATCAGACAGATTTTAACTCTGCGCATGTTGTTTGA 5729  
DB 2364 aatgatttaaaagttgttgaataaataattatgtatttggtttgggggtttagattt 2423  
QY 5730 TTTTGTGTTTGTAGCAGTGTGACTTAAGCCGAAGTTTGTAAAGTACATAAATCCAAATTT 5789  
DB 2424 tttgttgatttaaatattatttggtttggaagtggttattagattttttttaaaaaa 2483  
QY 5790 AATATGAAACCAACATTAATTTAAGTGTGAGAACTTATGCTTTAATTTGATTAATTTTG 5849  
DB 2484 tttgtttaaaaaatgatttggttttaattagtgatataataatcgatgaggtttt 2543  
QY 5850 TCGAGTATCATATGTGTGCAATTTGACTCAAAAATGAGTACCTTCAATTTAATTAATGATA 5909  
DB 2544 ttgaggaaagttaattatgttgagtaatgtatatttttggatttggatataattttgatt 2603  
QY 5910 TCTTCATACCAATGCTCCTTAAAGTGTTTGTAAAGATATCAATGCTTGAATAGACC 5969  
DB 2604 tataaataatgttttttattgatagtggttttgattgggaaaatttttgtaataatgatt 2663  
QY 5970 TAATTTGTAGACTTAAGACTTTTATTTTCTTAACCTTGATGCTTGCTTAAAGCATT 6029  
DB 2664 tgttttttttattttatttatttataaagtataattatgatttatttatttatt 2723  
QY 6030 TATCTATATCATATGATATGATGACGCCGCTGAGAACCAATTTCTGATTTTATATGTTTA 6089  
DB 2724 tataataatttgatgttgaagatttgggtttatatattttttttatttatttattt 2783  
QY 6090 TATTTCTTCTTAATGAACCTTAGAAAGACTACATGTTTACTAAGCAGGCCACTTTATGCT 6149  
DB 2784 tattttatgaacgaagaattatttggtttgagtaattgttgagaattatataatttaggt 2843  
QY 6150 TGTTTT 6155  
DB 2844 tttttt 2849

RESULT 30

AAV41256  
ID AAV41256 standard; cDNA; 2078 BP.

XX	AAV41256;	
AC		
XX		
DT	09-NOV-1998	(first entry)
XX		
DE	Human neuronal PAS domain protein NPAS1 cDNA.	
XX		
KW	NPAS1; neuronal bHLH-PAS domain; human; transcription factor;	
KW	therapy; diagnosis; ds.	
XX		
OS	Homo sapiens.	
XX		
FT	Key	Location/Qualifiers
FT	CDS	181..1953
FT		/*tag= a
XX		
PN	MO9831804-A1.	
XX		
PD	23-JUL-1998.	
XX		
PF	21-JAN-1998;	98KO-US01154.
XX		
PR	21-JAN-1997;	97US-0785310.
XX		
PA	(TEXA ) UNIV TEXAS SYSTEM.	
XX		
PI	McKnight SL, Russel DW;	
DR	WPI; 1998-414103/35.	
DR	P-PSDB; AAM68091.	
XX		
PT	New isolated neuronal PAS domain proteins - can regulate function of	
PT	neurological tissue such as brain tissue, used to develop products	
PT	for diagnosis and therapy	
XX		
PS	Claim 7; Page 21: 42pp; English.	
XX		
CC	This cDNA clone codes for human neuronal PAS domain protein NPAS1	
CC	(see AAM68091), a new member of the basic helix-loop-helix (bHLH)-PAS	
CC	family of transcription factors. A database search for expressed	
CC	sequence tags bearing sequence similarity to the PAS domain of the	
CC	aryl hydrocarbon receptor yielded 2 clones. These were used to	
CC	generate primers for PCR amplification of hybridisation probes,	
CC	(see AAM68091-94) coding for human and mouse NPAS proteins	
CC	(see AAM68091-94) were isolated from mouse brain tissue, human brain	
CC	tissue and HeLa cells. The human NPAS1 gene was mapped to	
CC	chromosome 19q13.2-q13.3. The NPAS proteins can regulate the	
CC	function of neurological tissue such as brain tissue. They can be	
CC	produced recombinantly from transformed host cells or purified from	
CC	mammalian cells. NPAS proteins and polynucleotides can be used in	
CC	diagnosis (e.g. genetic hybridisation screens for NPAS transcripts),	
CC	therapy (e.g. gene therapy to modulate NPAS gene expression) and in	
CC	the biopharmaceutical industry (e.g. as immunogens, reagents for	
CC	isolating other transcriptional regulators, and reagents for	
CC	screening chemical libraries for lead pharmacological agents).	
XX		
Q0	Sequence 2078 BP; 337 A; 751 C; 670 G; 320 T; 0 other;	

Query Match	0.98	Score 55.4	DB 19	Length 2078
Best Local Similarly	54.78	Pred. No. 0.00031		
Matches 110; Conservative	0	Mismatches 91	Indels 0	Gaps 0

**OY** 491 TCCACAGGATGCCCTGGGGCCTCATATGCTTAGGGCCCTGATGGTCTCTTTGTAG 550  
| | | | |  
**Db** 578 tcttcgcagcacccattgggaggatcactcttgcacgtccccttggaatgagctttgltgcct 637  
| | | | |  
**OY** 551 TGAACTGGAGGCACAAGTTGTGTTTGTCACAGAAATGTACACACTATCTAAGGATA 610  
| | | | |  
**Db** 638 tgaaaccggagaagaattctctcaactctccagsgaagcttcatactatctgygtctc 697  
| | | | |  
**OY** 611 ACCAGAAGACCTGATGAACAAAAGTATATTAGACTCTTGCAATGTTGGGGACCACACGG 670  
| | | | |

Db	698	cacgagtcgagatgaacgagcagcgcgtcttcgaatacaacctggggaacctcag	757
Qy	671	AATTTGTCAAAACCTGCTGC	691
Db	758	aggtctctgagcaactgggc	778

## RESULT 31

ID AAZ34629 standard; cDNA; 3486 BP.

AC AAZ34629;

DT 15-FEB-2000 (first entry)

Human clock cDNA.

KW CLOCK; transcription factor; circadian rhythm; human;

KW fertility; therapy; ss.

OS Homo sapiens.

FH	Key	Location/Qualifiers
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9
10	10	10
11	11	11
12	12	12
13	13	13
14	14	14
15	15	15
16	16	16
17	17	17
18	18	18
19	19	19
20	20	20
21	21	21
22	22	22
23	23	23
24	24	24
25	25	25
26	26	26
27	27	27
28	28	28
29	29	29
30	30	30
31	31	31
32	32	32
33	33	33
34	34	34
35	35	35
36	36	36
37	37	37
38	38	38
39	39	39
40	40	40
41	41	41
42	42	42
43	43	43
44	44	44
45	45	45
46	46	46
47	47	47
48	48	48
49	49	49
50	50	50
51	51	51
52	52	52
53	53	53
54	54	54
55	55	55
56	56	56
57	57	57
58	58	58
59	59	59
60	60	60
61	61	61
62	62	62
63	63	63
64	64	64
65	65	65
66	66	66
67	67	67
68	68	68
69	69	69
70	70	70
71	71	71
72	72	72
73	73	73
74	74	74
75	75	75
76	76	76
77	77	77
78	78	78
79	79	79
80	80	80
81	81	81
82	82	82
83	83	83
84	84	84
85	85	85
86	86	86
87	87	87
88	88	88
89	89	89
90	90	90
91	91	91
92	92	92
93	93	93
94	94	94
95	95	95
96	96	96
97	97	97
98	98	98
99	99	99
100	100	100

```

ET      / *tag= a

```

PN W09957137-A1.

PD 11-NOV-1999.

PF 06-MAY-1999; 99WO-US10072.

PR 07-MAY-1998; 98US-0084610.

PA (HARD ) HARVARD COLLEGE.

PI Weitz CJ, Gekakis N, Staknis D;

WPI: 2000-052938/04.

XX

# PT diagnosing and treating circadian clock disruption disorders -

PS Disclosure; Fig 25A-B; 96pp; English.

CC This is the nucleotide sequence of human CDNA encoding CLOCK

CC per1 gene when present in combination with BMAL1 (see AAY32209).

CC mechanism regulating genes responsible for the establishment and/or

CC novel drugs aimed at restoration of a normal circadian cycle, the

CC E-box-linked genes. The drugs are used to treat conditions such as

CC and infertility. The invention also provides BMAL1 and CLOCK

CC E-box-linked gene which regulates the circadian clock.

Sequence 3486 BP; 1097 A; 755 C; 721 G; 913 T; 0 other;

Query Match	0.98;	Score 55;	DB 21;	Length 3486;
Best Local Similarity	55.58;	Pred. No. 0.00057;		
Matches 106; Conservative	0;	Mismatches 85;	Indels 0;	Gaps 0;

QY 456 AAGAGTGGCGTGGGGCCATATATCTGTAGAGCCCTTGATGGGTCTTCTTTGATGTAAC 5555

Db 676 aatgaagaggtttacaacattaatgtttgaagagctctctatggttttttttttttagcattatg 7555

QY 556 CTGAGAGCAACGTTGTGTGTGTGTCTGACAGAGATGTGACACAGTCTTCAAGGTATTAACCA 6155

Db 736 acagatggaagcatalaatactgtctcgagagtgtaacttcactgaacattacca 795  
 QY 616 GAAGAGCTGATGAACAAGTGTATATAGCATCTTGGGACACACAGGAATT 675  
 Db 796 tctgattctgtgatacaagatatattatccaccagaagggaacattcagaggt 855  
 QY 676 GTCAAAAACCT 686  
 Db 856 tataaataact 866  
 RESULT 32  
 AAV61450  
 ID AAV61450 standard; DNA: 3546 BP.  
 AC AAV61450;  
 XX  
 DT 02-FEB-1999 (first entry)  
 DE Human Clock gene.  
 XX  
 KM Clock gene: circadian rhythm; human; jet lag; sleep-wake disorder;  
 XX seasonal affective disorder; cancer; diagnosis; therapy; ds.  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 418..2958  
 FT /\*tag= a  
 XX  
 PN WO9840514-A1.  
 PD 17-SEP-1998.  
 XX  
 PF 13-MAR-1998; 98MO-US05114.  
 XX  
 PR 30-JUN-1997; 97US-0885291.  
 PR 13-MAR-1997; 97US-0816693.  
 XX  
 PA (NOUN ) UNIV NORTHWESTERN.  
 XX  
 PI Pinto LH, Takahashi JS, Turek F;  
 XX  
 DR WPI: 1998-520828/44.  
 DR P-PSDB; AAW79533.  
 XX  
 PT Novel mouse and human circadian rhythm gene, clock - useful for  
 PT treating e.g. jet-lag, sleep-wake disorders, abnormal cell division,  
 PT etc  
 PS Claim 5; Fig 14.1-14.3; 154pp; English.  
 XX  
 CC This is the nucleotide sequence of the human Clock gene that codes  
 CC for a putative 846-amino acid polypeptide (see AAW79533) involved in  
 CC the regulation of circadian rhythm. To isolate the gene, several  
 CC human clones identified in the NCBI database by end-sequence  
 CC analysis were sequenced, and a human hypothalamus cDNA library was  
 CC screened to isolate novel clones that hybridised with a probe of  
 CC the mouse Clock gene (see AAV61401). Further DNA sequence alignments  
 CC to the transcript of the mouse Clock gene revealed that a consensus  
 CC sequence from the aggregate of EST and hypothalamic clones extended  
 CC through the gene's entire coding region and into much of its  
 CC flanking 5' and 3' untranslated regions. The Clock gene regulates  
 CC at least 2 fundamental properties of the circadian clock system:  
 CC the intrinsic circadian period and the persistence of circadian  
 CC rhythmicity. The invention provides isolated and purified Clock  
 CC polypeptides, polynucleotides (including antisense), vectors and  
 CC host cells. These can be used to treat disorders of altered or  
 CC disrupted circadian rhythms e.g. jet-lag, seasonal affective  
 CC disorder, sleep-wake cycle disorders such as mood state, stress,  
 CC neurological disorders, to regulate diet and food intake especially  
 CC for diabetes, to treat cardiovascular, respiratory, liver or

CC endocrine disorders, and for diagnosis and treatment of abnormal  
 CC cell division such as cancer.  
 XX  
 SQ Sequence 3546 BP; 1118 A; 772 C; 728 G; 928 T; 0 other;  
 Query Match 0.9%; Score 55; DB 19; Length 3546;  
 Best Local Similarity 55.5%; Pred. No. 0.00058;  
 Matches 106; Conservative 0; Mismatches 85; Indels 0; Gaps 0;  
 QY 496 AAGATGCGCTGGGCGCTATGATGCTTGAGCCCTTGATGGTCTTTGTAGTGAA 555  
 Db 736 aatgaagagttacacataatagtagagccttgatgttcttttaagaacatcat 795  
 QY 556 CTGGAAGCAACGTTGTGTTGTCTCAGAGATGTGACACAGATCTTAAGGTATTA 615  
 Db 796 acagatggaagcatalaatactgtctcgagagtgtaacttcactgaacattacca 855  
 QY 616 GAAGAGCTGATGAACAAGTGTATATAGCATCTTGGGACACACAGGAATT 675  
 Db 856 tctgattctgtgatacaagatatattatccaccagaagggaacattcagaggt 915  
 QY 676 GTCAAAAACCT 686  
 Db 916 tataaataact 926  
 RESULT 33  
 AAX03456  
 ID AAX03456 standard; cDNA: 5715 BP.  
 XX  
 AC AAX03456;  
 XX  
 DT 10-MAY-1999 (first entry)  
 DE Human HSCLOCK cDNA.  
 XX  
 KM HSCLOCK; clock gene; circadian rhythm; body clock; human;  
 XX sleep disorder; jet lag; diagnosis; therapy; ss.  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 252..2792  
 FT /\*tag= a  
 XX  
 PN EP890578-A2.  
 PD 13-JAN-1999.  
 XX  
 PF 26-JUN-1998; 98BP-0305065.  
 XX  
 PR 08-JUL-1997; 97EP-0304996.  
 XX  
 PA (SMIR ) SMITHKLINE BEECHAM PLC.  
 XX  
 DR Duckworth DM, Michalovich D;  
 DR WPI: 1999-072834/07.  
 DR P-PSDB; AAW84565.  
 XX  
 PT New uses of clock gene (HSCLOCK) polypeptides and polynucleotides -  
 PT useful as diagnostic reagents and for treatment of sleep disorders,  
 PT jet lag and old age pathologies  
 PS Disclosure; Page 11-13; 22pp; English.  
 XX  
 CC This polynucleotide comprises human HSCLOCK cDNA. The cDNA can be  
 CC isolated from a human brain cDNA library by expressed sequence tag  
 CC analysis, and can be used for producing HSCLOCK polypeptide (see  
 CC AAW83465) by recombinant methods. Sequence homology to the mouse  
 CC clock gene is about 87%. Recombinant host cells comprising an  
 CC expression system capable of producing a clock gene (HSCLOCK)



```

FT      /tag= a
XX      /partial
PN      WO928464-A2.
XX      10-JUN-1999.
PD      27-NOV-1998; 98WO-US25314.
XX      28-NOV-1997; 97US-0066863.
XX      (WISC ) WISCONSIN ALUMNI RES FOUND.
PA      Bradfield CA, Gu YZ, Hogenesch JB;
PI      WPI: 1999-371120/31.
XX      P-PSDB; AAY06293.
XX      Developmental signal transduction associated proteins
PS      Claim 6; Page 97-98; 106pp; English.
XX      This is the nucleotide sequence of a partial cDNA encoding MOP5 (see
CC      AAY06293), a novel member of the PAS superfamily, where PAS stands
CC      for PER/ARNT/SIM domains. The cDNA was identified in an iterative
CC      search of human ESTs designed to identify basic-helix-loop-helix-PAS
CC      (bHLH-PAS) proteins that interact with either the Ah receptor (AHR)
CC      or the Ah receptor nuclear translocator (ARNT). To obtain extended
CC      open reading frames for each EST, an anchored-PCR strategy was used
CC      to amplify additional flanking sequences from a commercial HepG2
CC      cDNA library. The invention provides novel MOP5 2-9 nucleic acids
CC      (see AAY58881-88) and proteins (see AAY06289-97). These are useful in a
CC      variety of research, diagnostic and therapeutic applications.
CC      Several of the MOP5 are alpha-class hypoxia-inducible factors.
CC      Others are involved in circadian signal transduction.
XX      Sequence 1581 BP; 276 A; 560 C; 494 G; 251 T; 0 other;
SQ
Query Match      0.9%; Score 54.2; DB 20; Length 1581;
Best Local Similarity 55.6%; Pred. No. 0.00055;
Matches 104; Conservative 0; Mismatches 83; Indels 0; Gaps 0;
QY      505 CTGGGGCCATATGATGCTTGAGGCCCTTGATGGTCTCTTTGTAGTGAACCTGGAAGC 564
DB      11111 1 11 111111111 11 111111111
DB      86 ctggggggtacatcttcgagacagtcctcgtggttggctgccttgacaacgaaga 145
QY      565 AACGTGTGTTTGTGACGAGATGTGACACATATCTAAGTATACCAAGAAGAGCTG 624
DB      11 11 1111111 11 11111 11111 11 11111
DB      146 aaatcctctacatctcagagacagtcctcattatcttcctcagagtgagatg 205
QY      625 ATGAAACAAAAGTGTATATAGCATCTTGACATGTTGGGACACACAGAAATTGTCAAAAAC 684
DB      11 11 11111 11 111111111 11 111111111
DB      206 acggcgacagcagcgtctcgtactacatcaccctgggaccactcagagtgctgagca 265
QY      685 CTGCTGC 691
DB      111 11
DB      266 ctggggc 272
RESULT 36
AAFI8177
ID      AAFI8177 standard; DNA; 4260 BP.
XX
AC      AAFI8177;
XX
DT      14-MAR-2001 (first entry)
XX
DE      Lung cancer associated polynucleotide sequence SEQ ID 196.
XX
KW      Human; lung cancer associated protein; neuroprotective; cytostatic;
KW      cardioactive; immunomodulatory; muscular active; vulnary;
KW      gastrointestinal; nephrotoxic; antiinfective; gynecological;

```

```

KW      antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
KW      proliferative disorder; wound healing; infectious disease; ds.
XX
OS      Homo sapiens.
XX      WO200055180-A2.
XX      21-SEP-2000.
XX      08-MAR-2000; 2000WO-US05918.
XX      12-MAR-1999; 99US-0124270.
XX      (HUMA-) HUMAN GENOME SCI INC.
PA      (ROSE/) ROSEN C A.
XX      Ruben SM;
XX      WPI: 2000-587514/55.
XX      P-PSDB; AAB58301.
XX      Lung cancer associated gene sequences, referred to as lung cancer
PT      antigens, useful for treatment, prevention, and diagnosis of disorders
PT      such as lung cancer -
XX      Claim 1; Page 661-663; 1425pp; English.
XX      Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
CC      associated proteins represented in AAB58106 - AAB55548. Lung cancer
CC      associated proteins and polynucleotide sequences, their agonists, and
CC      antagonists may have neuroprotective; cytostatic; cardioactive;
CC      immunomodulatory; muscular active general; vulnary; gastrointestinal
CC      general; nephrotoxic; antiinfective; gynecological; or antibacterial
CC      activity. The invention also includes antibodies specific for the
CC      protein or polynucleotide sequences. The lung cancer associated
CC      polynucleotide sequences may be used for detection of lung cancer,
CC      chromosome identification, as chromosome markers, and for numerous other
CC      diagnostic or research purposes. The proteins may be used to treat
CC      disorders such as neural, immune, muscular, reproductive,
CC      gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC      disorders. The proteins may also be used in the treatment of wounds and
CC      infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and
CC      peptide AAB58349 are used in the course of the invention for the
CC      identification and characterisation of the polynucleotide and protein
CC      sequences.
XX      Sequence 4260 BP; 1297 A; 993 C; 887 G; 1077 T; 6 other;
SQ
Query Match      0.9%; Score 53.8; DB 21; Length 4260;
Best Local Similarity 52.0%; Pred. No. 0.0014;
Matches 250; Conservative 0; Mismatches 212; Indels 19; Gaps 5;
QY      4068 TCCAACTACGGAATAAGTACGACACCTGATCCAGGCTTTACTGGGGCTACGACTGCCCA 4127
DB      11111 11111 11 111 111 11111 11 11111
DB      1529 tccagagcaggaatggttccccaagtggagccaacttgcctcattcctaagccctg 1588
QY      4128 GAGCCCACTTATGTCACCCCGAATGGCAGCATACAGAGCCCATGATGCAACAGTCTCA 4187
DB      11111 11 11 11111 11 1111111 11 11 11 111
DB      1589 gagctcattggtgcgattgcacatccctcctcagagttctctcctcagcaaac 1648
QY      4188 GGGCAACCCAGCCTATCAGGCCCTTCGACATAAATGATGGGCGCAGGGAATGGG 4247
DB      11 11 1111111 11 11111 11 11111 11
DB      1649 acctgcctccgggtatcagtcaccca--gacatgaagcctggcagcaagagcgata-- 1703
QY      4248 CGAAGACGATGTTTCCAGCAGTCCGACACCATTTGGGCGACGAACACACAG 4307
DB      1111111 11 11 11 11 11 11 11 11 11 11 11
DB      1704 -ggaaacaacaatggttccagtcacagctgtccagacagccacgcctgcacagcag 1762
QY      4308 CATGTACAGTAAACATGAAATGATGTCATGTCATGGCAGCAACACAGAGTGAG 4367
DB      11 111 11111111 11111 11 111111 111111 11
DB      1763 agtatcc--aacacatgagcatcacggttccatcag-----cagtggaataac 1810

```



[illegible]





**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 12, 2002, 03:38:46 ; Search time 10727.9 Seconds  
(without alignments)  
12008.337 Million cell updates/sec

Title: US-09-842-256-1  
Perfect score: 6156  
Sequence: 1 GCGCGCCGACGCTCGGCTA.....CCACTTTATGTTGTTT 6156

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues  
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_hgt:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_hgt\_hum:\*  
31: em\_hgt\_inv:\*  
32: em\_hgt\_other:\*  
33: em\_hgtgo\_inv:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No. Query Match length DB ID Description

1	6154.4	100.0	6156	9	HSTIF2GEN	X97674 H.sapiens m
2	3846	62.5	4878	10	MMU39060	U39060 Mus musculus
3	3699.6	60.1	4771	10	AF0005082	AF0005082 Mus muscu
4	3593.2	58.4	4398	10	AF136943	AF136943 Rattus no
5	1961.4	31.9	4871	5	X1A243119	A1243119 Xenopus l
6	1592.8	25.9	86554	2	AC090280	AC090280 Homo sapi
7	1592.8	25.9	181403	2	AC084251	AC084251 Homo sapi
8	1591.2	25.8	71701	2	AF254984	AF254984 Homo sapi
9	1274.2	20.7	86554	2	AC090280	AC090280 Homo sapi
10	1274.2	20.7	122924	2	AF207954	AF207954 Homo sapi
11	1274.2	20.7	124630	9	AC090731	AC090731 Homo sapi
12	1238.4	20.1	4958	5	AF323989	AF323989 Danio rer
13	1071.2	17.4	110000	2	AC091288-0	AC091288 Mus muscu
14	1036	16.8	155871	2	AC095150	AC095150 Rattus no
15	564.2	9.2	4609	10	AF000581	AF000581 Mus muscu
16	553.4	9.0	4495	9	AF010227	AF010227 Homo sapi
17	542.8	8.8	6835	9	AF012108	AF012108 Homo sapi
18	533.6	8.7	4668	9	AF016031	AF016031 Homo sapi
19	524.6	8.5	110000	2	AC091288-0	AC091288 Mus muscu
20	524.6	8.5	223169	2	AC091248	AC091248 Mus muscu
21	517.6	8.3	4546	5	AF044080	AF044080 Xenopus l
22	510.6	8.3	6754	9	AF036892	AF036892 Homo sapi
23	454.8	7.4	110000	2	AC091288-1	Continuation (2 of
24	369.8	6.0	71701	2	AF254984	AF254984 Homo sapi
25	369.6	6.0	5169	10	MMU64828	U64828 Mus musculu
26	368	6.0	4752	10	MMU64606	U64606 Mus musculu
27	366.4	6.0	4752	10	MMU64606	U64606 Mus musculu
28	366.2	5.9	4664	9	HSJ000881	AJ000881 Homo sapi
29	366.2	5.9	4721	9	HSJ000882	AJ000882 Homo sapi
30	364.6	5.9	4660	9	HSU59302	U59302 Human stero
31	363.2	5.9	4547	9	HSU90661	U90661 Human stero
32	362.8	5.9	4527	5	AF443296	AF443296 Coturnix
33	339.8	5.5	399	6	AX351515	AX351515 Sequence
34	250	4.1	250	11	G15398	G15398 human STS S
35	237.4	3.9	104593	2	AC091288-3	Continuation (4 of
36	195.4	3.2	4036	10	AF322224	AF322224 Rattus no
37	194	3.2	155871	2	AC095150	AC095150 Rattus no
38	192	3.1	70606	2	AC109202	AC109202 Mus muscu
39	154.4	2.5	223169	2	AC091248	AC091248 Mus muscu
40	150.6	2.4	1367	9	AK021838	AK021838 Homo sapi
41	136.6	2.2	110000	2	AC091288-1	Continuation (2 of
42	132.2	2.1	207810	2	AC091158	AC091158 Mus muscu
43	102.8	1.7	86453	9	HS1049316	AL034418 Human DNA
44	101.4	1.6	188741	2	AC013459	AC013459 Homo sapi
45	95.4	1.5	170936	10	AL589873	AL589873 Mouse DNA

## ALIGNMENTS

RESULT 1  
HSTIF2GEN LOCUS 6156 bp mRNA linear PRI 10-MAR-1997  
DEFINITION H.sapiens mRNA for transcriptional intermediary factor 2.  
X97674.1  
VERSION X97674.1 GI:1877214  
KEYWORDS alternatively spliced; nuclear receptor coactivator; Tif2 gene; transcriptional mediator.  
SOURCE  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 6156)  
AUTHORS Voegel,J.-J., Heine,M.J., Zechel,C., Chambon,P. and Gronemeyer,H.  
TITLE TIF2, a 160 kDa transcriptional mediator for the ligand-dependent activation function AF-2 of nuclear receptors  
JOURNML EMBO J. 15 (14), 3667-3675 (1996)  
MEDLINE 96312964  
REFERENCE 2 (bases 1 to 6156)  
AUTHORS Voegel,J.-J.  
TITLE Direct Submission  
JOURNAL Submitted (22-APR-1996) J.J. Voegel, IGBMC Inst.de Genet.et Biol.Mol.et Cell., CNRS-INSERM-Univ.Louis Pasteur, B.P.163, C.U. de



QY 1321 ACTGACAAAGCATGGGAGACCACTGAATCCATTAGCTCTACAGCCCTGCCCATCAG 1380  
|||||  
Db 1321 ACTGACAAAGCATGGGAGACCACTGAATCCATTAGCTCTACAGCCCTGCCCATCAG 1380  
QY 1381 GCCCTGTGCACTGGGAGACCCAGGTAGACATGACCTCTCAGTACGAATTTAAATTTTCCC 1440  
|||||  
Db 1381 GCCCTGTGCACTGGGAGACCCAGGTAGACATGACCTCTCAGTACGAATTTAAATTTTCCC 1440  
QY 1441 ATAAATGGCCCAAGAGAAATGGGCAATGGCCATGGGAGGTTGGTGGTCTGGGGGA 1500  
|||||  
Db 1441 ATAAATGGCCCAAGAGAAATGGGCAATGGCCATGGGAGGTTGGTGGTCTGGGGGA 1500  
QY 1501 ATGAACCATGTGTGAGGATGCAAGCAACACTCTCTCAGGGTACTAGTATGCACTCAAA 1560  
|||||  
Db 1501 ATGAACCATGTGTGAGGATGCAAGCAACACTCTCTCAGGGTACTAGTATGCACTCAAA 1560  
QY 1561 ATGAACAGCCCTCTCAAAAGCAGCCCTGGCATGAATCCAGACAGCCCACTCCATGCTT 1620  
|||||  
Db 1561 ATGAACAGCCCTCTCAAAAGCAGCCCTGGCATGAATCCAGACAGCCCACTCCATGCTT 1620  
QY 1621 TCACCAAGCATGCGATGAGGCTGGAGTGGCTGGACGCTCGAATGCCACCCATGCTAG 1680  
|||||  
Db 1621 TCACCAAGCATGCGATGAGGCTGGAGTGGCTGGACGCTCGAATGCCACCCATGCTAG 1680  
QY 1681 TTTTCCCTGAGAGAAAGCTTCATTCCTCTGTGGAGTTTGCAGACAGACAGAAATAGC 1740  
|||||  
Db 1681 TTTTCCCTGAGAGAAAGCTTCATTCCTCTGTGGAGTTTGCAGACAGACAGAAATAGC 1740  
QY 1741 CATAGTTATACCAACAGCTCCCTCAATGCACTTCAGGCCCTTCAGGAGGAGGACGGGCTC 1800  
|||||  
Db 1741 CATAGTTATACCAACAGCTCCCTCAATGCACTTCAGGCCCTTCAGGAGGAGGACGGGCTC 1800  
QY 1801 TCATTAGGGTATGCTGGTGGCTTCACACAGACTTAAAAATGGGCAATTTGGCAAACTCCCCA 1860  
|||||  
Db 1801 TCATTAGGGTATGCTGGTGGCTTCACACAGACTTAAAAATGGGCAATTTGGCAAACTCCCCA 1860  
QY 1861 GTTAATATGAATCCCTCCCACTCAGCAAGATGGAGACTTGGACTCAAAAGACTGTTT 1920  
|||||  
Db 1861 GTTAATATGAATCCCTCCCACTCAGCAAGATGGAGACTTGGACTCAAAAGACTGTTT 1920  
QY 1921 GGACTATATGGGGAGCCCTCTGAAAGTACAACTGGACAAGCAGAGAGAGCTGCCATCT 1980  
|||||  
Db 1921 GGACTATATGGGGAGCCCTCTGAAAGTACAACTGGACAAGCAGAGAGAGCTGCCATCT 1980  
QY 1981 GGAGAGCAAAAGAGAAACAATGACCCCACTGCCCGCGCGCTGAGAGAGAGAGACT 2040  
|||||  
Db 1981 GGAGAGCAAAAGAGAAACAATGACCCCACTGCCCGCGCGCTGAGAGAGAGAGACT 2040  
QY 2041 GACGGGCGAGCAGACTGTCATGACAGCAAAAGGGCAGACCAAACTCTCAGCTGTGACC 2100  
|||||  
Db 2041 GACGGGCGAGCAGACTGTCATGACAGCAAAAGGGCAGACCAAACTCTCTCAGCTGTGACC 2100  
QY 2101 ACCAAATCTGATCAGATGAGACCCCTGCGCTTAGCCAGACTCTTGTGCGATACAAACAA 2160  
|||||  
Db 2101 ACCAAATCTGATCAGATGAGACCCCTGCGCTTAGCCAGACTCTTGTGCGATACAAACAA 2160  
QY 2161 GACTCCACAGTAGCTTCCCTGGTCTCTGGGTCTACACATGGAACCTCCCTCAAGAGAG 2220  
|||||  
Db 2161 GACTCCACAGTAGCTTCCCTGGTCTCTGGGTCTACACATGGAACCTCCCTCAAGAGAG 2220  
QY 2221 CATAAATATTTTGCACAGACTTTCGAGAGACAGTTCCTCTGTGAGATTGGCAAGTTA 2280  
|||||  
Db 2221 CATAAATATTTTGCACAGACTTTCGAGAGACAGTTCCTCTGTGAGATTGGCAAGTTA 2280  
QY 2281 ACACAGAAAGCCACAGGCAAAAGACTCTGAGCCAGAGATCCAGACAGCTCTCTGATCA 2340  
|||||  
Db 2281 ACACAGAAAGCCACAGGCAAAAGACTCTGAGCCAGAGATCCAGACAGCTCTCTGATCA 2340  
QY 2341 GAAGTACTATTTAAACAAGAGCGGTGAGCCCAAGAAAGAAAGAAATGCACTACTTGGC 2400  
|||||  
Db 2341 GAAGTACTATTTAAACAAGAGCGGTGAGCCCAAGAAAGAAAGAAATGCACTACTTGGC 2400  
QY 2401 TATTTGTAGATAAAGATGATTAAGATATTTGGTTTACGAGAAATTAACCCCAACTT 2460  
|||||

Db 2401 TATTTGTAGATAAAGATGATTAAGATATTTGGTTTACGAGAAATTAACCCCAACTT 2460  
QY 2461 GAGACACTGGACAGTAAAGACACATCTCCAGTAAACAAAAATTAATGCAATGAAACT 2520  
|||||  
Db 2461 GAGACACTGGACAGTAAAGACACATCTCCAGTAAACAAAAATTAATGCAATGAAACT 2520  
QY 2521 GAGAAGGAGAGATGAGCTTGTAGCCTGGTGTGACCGCCGGAGAGAGCTGAGCAACTTG 2580  
|||||  
Db 2521 GAGAAGGAGAGATGAGCTTGTAGCCTGGTGTGACCGCCGGAGAGAGCTGAGCAACTTG 2580  
QY 2581 GAGAGATTTTGGATGATTTTGCAGAAATAGTCAATTAACACAGCTTTTCCACAGACGAG 2640  
|||||  
Db 2581 GAGAGATTTTGGATGATTTTGCAGAAATAGTCAATTAACACAGCTTTTCCACAGACGAG 2640  
QY 2641 CCAGCGCCCTGCTGTGATCACTTACAAAGCAAGCCATCATTAATGACCTCATGCACTC 2700  
|||||  
Db 2641 CCAGCGCCCTGCTGTGATCACTTACAAAGCAAGCCATCATTAATGACCTCATGCACTC 2700  
QY 2701 ACAGCTGAAAACAGCCCTGTACACCTGTGTGGAGCCCAAGAAACAGCACTGGGAATTTCA 2760  
|||||  
Db 2701 ACAGCTGAAAACAGCCCTGTGTACACCTGTGTGGAGCCCAAGAAACAGCACTGGGAATTTCA 2760  
QY 2761 CAGACACTTTTAAATTAACCCACAGACGAGGCAACTGGGCAAGTTATGCAAAACAGAAAT 2820  
|||||  
Db 2761 CAGACACTTTTAAATTAACCCACAGACGAGGCAACTGGGCAAGTTATGCAAAACAGAAAT 2820  
QY 2821 TTACCACTTGACATCATTTGCAAAAGCCCACTGTGCTGACCTTTCCACCAATTCAGA 2880  
|||||  
Db 2821 TTACCACTTGACATCATTTGCAAAAGCCCACTGTGCTGACCTTTCCACCAATTCAGA 2880  
QY 2881 AACACTAGTCCCTACTCATGATTAACCTCAGCCAGGAATGATGGATTAATCAAGGATGATA 2940  
|||||  
Db 2881 AACACTAGTCCCTACTCATGATTAACCTCAGCCAGGAATGATGGATTAATCAAGGATGATA 2940  
QY 2941 GGAACCAAGGAAATTTAGGGAACAGTAGACAGGAATGATTTGTAACAGTCTCTCGG 3000  
|||||  
Db 2941 GGAACCAAGGAAATTTAGGGAACAGTAGACAGGAATGATTTGTAACAGTCTCTCGG 3000  
QY 3001 CCTACTATGCCATCTGAGAAATGGGACCGCAGAGTGGGCTGTGAGAGTCACTGTGCT 3060  
|||||  
Db 3001 CCTACTATGCCATCTGAGAAATGGGACCGCAGAGTGGGCTGTGAGAGTCACTGTGCT 3060  
QY 3061 GCTACCAAGAGTGCATGAACCGGCACTGCAAGGAGATGATTCGGAACCCACAGCC 3120  
|||||  
Db 3061 GCTACCAAGAGTGCATGAACCGGCACTGCAAGGAGATGATTCGGAACCCACAGCC 3120  
QY 3121 AGCATCCCATGAGGCGCCAGCAGCAGCCTGGCCAAAGACAGACGCTTCAGTCAAGGTC 3180  
|||||  
Db 3121 AGCATCCCATGAGGCGCCAGCAGCAGCCTGGCCAAAGACAGACGCTTCAGTCAAGGTC 3180  
QY 3181 ATGAATATAGGGCCATCTGAAATTAAGATGAATGAGGGGGGACCTCAGATATGCCAACA 3240  
|||||  
Db 3181 ATGAATATAGGGCCATCTGAAATTAAGATGAATGAGGGGGGACCTCAGATATGCCAACA 3240  
QY 3241 CAAGCTCCCAAAATCAGAGTCCCGCATGGGCTGAAAGATCTCGCTATAGACAGGCG 3300  
|||||  
Db 3241 CAAGCTCCCAAAATCAGAGTCCCGCATGGGCTGAAAGATCTCGCTATAGACAGGCG 3300  
QY 3301 TCTTTTGCAGGCAAAACAGGAGCCATTTGGCACTTTCAGATGATGCTATGTGCA 3360  
|||||  
Db 3301 TCTTTTGCAGGCAAAACAGGAGCCATTTGGCACTTTCAGATGATGCTATGTGCA 3360  
QY 3361 CATCTCGAGCTGAGTCTCCAGATGATAGGAGAGTCTCTGAGCAGACCTGTATGTGCC 3420  
|||||  
Db 3361 CATCTCGAGCTGAGTCTCCAGATGATAGGAGAGTCTCTGAGCAGACCTGTATGTGCC 3420  
QY 3421 TTGCGGAATTTTGAATGGGCTGGAGAGATGATTAAGGCTTAAGAAATCCGAATGTGTC 3480  
|||||  
Db 3421 TTGCGGAATTTTGAATGGGCTGGAGAGATGATTAAGGCTTAAGAAATCCGAATGTGTC 3480  
QY 3481 AGCCAGAGCCAAAGAGTATGATCCAGAACAGTTCTCAAGTCAAGATTCACATCTGCTG 3540  
|||||

Db 3481 AGCCAGAGCCAGAGATAGATCCAGAAACGTTCTCAAGTCAGCATTCCTCAACATCATCTG 3540  
QY 3541 GAGCAGAGAGCGCCGCTTTTCCACAGCAGATATGATCTGACAGCAAAATGGCCAGAGT 3600  
Db 3541 GAGCAGAGAGCGCCGCTTTTCCACAGCAGATATGATCTGACAGCAAAATGGCCAGAGT 3600  
QY 3601 AGCATTTCTCCATGACAGATTCCAAACCTTTTCACACATGGACAGCGGCTAGTTATGCC 3660  
Db 3601 AGCATTTCTCCATGACAGATTCCAAACCTTTTCACACATGGACAGCGGCTAGTTATGCC 3660  
QY 3661 ACATCTCGTATGAGCGCCAGCGGGCTGACGCGCCACGCGGCTAGTGGACCAACAGCA 3720  
Db 3661 ACATCTCGTATGAGCGCCAGCGGGCTGACGCGCCACGCGGCTAGTGGACCAACAGCA 3720  
QY 3721 AATCAACTAAGACTTCACTTACAGCATGCGCTCCAGCAGCAGAGAAATGCGCACCACTT 3780  
Db 3721 AATCAACTAAGACTTCACTTACAGCATGCGCTCCAGCAGCAGAGAAATGCGCACCACTT 3780  
QY 3781 ATGATATCAAAATCAGCAATGTTTCCAAATGTGAATTTGACTCTGAGGCTGGAGTACCA 3840  
Db 3781 ATGATATCAAAATCAGCAATGTTTCCAAATGTGAATTTGACTCTGAGGCTGGAGTACCA 3840  
QY 3841 CAGGACCTTATTAATGACAGATGCTGGCCAGAGACAGAGGAAATCCTGAACACAGAT 3900  
Db 3841 CAGGACCTTATTAATGACAGATGCTGGCCAGAGACAGAGGAAATCCTGAACACAGAT 3900  
QY 3901 CTTGACAGAGACAAATGATCAGACACAGCAAGTTGAGCAACGAACTTTGATGATGAGA 3960  
Db 3901 CTTGACAGAGACAAATGATCAGACACAGCAAGTTGAGCAACGAACTTTGATGATGAGA 3960  
QY 3961 GGCACAAAGGTTGAAATATGACACCAAGATGCTGCTAGTGGTATGCCAGCAACTATG 4020  
Db 3961 GGCACAAAGGTTGAAATATGACACCAAGATGCTGCTAGTGGTATGCCAGCAACTATG 4020  
QY 4021 AGCAACCTGCGATTTCCCGAGGCAATGCAAGCAGTTCATTTCCCTCCAAATACAGGA 4080  
Db 4021 AGCAACCTGCGATTTCCCGAGGCAATGCAAGCAGTTCATTTCCCTCCAAATACAGGA 4080  
QY 4081 ATAAATGACAACTGATTCACAGGCTTTACTGGGGCTACGACTCCCGAGGCCACTTATG 4140  
Db 4081 ATAAATGACAACTGATTCACAGGCTTTACTGGGGCTACGACTCCCGAGGCCACTTATG 4140  
QY 4141 TCAACCCGATGGGACATTCACAGAGTCCATGATGCAACAGTCTCAGGCCAACCCAGCC 4200  
Db 4141 TCAACCCGATGGGACATTCACAGAGTCCATGATGCAACAGTCTCAGGCCAACCCAGCC 4200  
QY 4201 TATCAGGCCCCCTCCGACATTAATGATGGGCGAGGGGAAATGGGCGGAAACAGCATG 4260  
Db 4201 TATCAGGCCCCCTCCGACATTAATGATGGGCGAGGGGAAATGGGCGGAAACAGCATG 4260  
QY 4261 TTTTCCAGCAGATGCCACACACTTTGGGCGAGCAACCAACACACATGATGATGATAC 4320  
Db 4261 TTTTCCAGCAGATGCCACACACTTTGGGCGAGCAACCAACACACATGATGATGATAC 4320  
QY 4321 AATCATGAACATGATGTCATGAGGAGCAACCAAGTGGCATGAGCAGATCAACAG 4380  
Db 4321 AATCATGAACATGATGTCATGAGGAGCAACCAAGTGGCATGAGCAGATCAACAG 4380  
QY 4381 ATGACAGAGAGATCAGATGACCTCAGTACCTGCTGTACGTAGAGGCTCTCTCC 4440  
Db 4381 ATGACAGAGAGATCAGATGACCTCAGTACCTGCTGTACGTAGAGGCTCTCTCC 4440  
QY 4441 ATGGGTCGCCAGCAGGTTAATGATCTGCTGAGGGAGGCAACCTGTTCCCAACAG 4500  
Db 4441 ATGGGTCGCCAGCAGGTTAATGATCTGCTGAGGGAGGCAACCTGTTCCCAACAG 4500  
QY 4501 CTGCTCTGGAATGATATGATTTAAGCAGAGGAGAGACAAACGGAATATGCTGACAC 4560  
Db 4501 CTGCTCTGGAATGATATGATTTAAGCAGAGGAGAGACAAACGGAATATGCTGACAC 4560  
QY 4561 TGGTGAAGCCAGTTGCTTCTTCACTGACGAGCGGGCTCACTTGTCTCAAAACCTTCCAGTCT 4620  
Db 4561 TGGTGAAGCCAGTTGCTTCTTCACTGACGAGCGGGCTCACTTGTCTCAAAACCTTCCAGTCT 4620

QY 4621 GGAGAGCTGTCTATTTGTTTCAACCCAACTGACCTGACGCGGTTGCTAGACAG 4680  
Db 4621 GGAGAGCTGTCTATTTGTTTCAACCCAACTGACCTGACGCGGTTGCTAGACAG 4680  
QY 4681 ACAGGCTGGCCCTGGTTCCAGAGGCGGCTCCACTGGCTGTGGCAGAGAGAGCTGCT 4740  
Db 4681 ACAGGCTGGCCCTGGTTCCAGAGGCGGCTCCACTGGCTGTGGCAGAGAGAGCTGCT 4740  
QY 4741 CTTCTCTGACAGCTGAGGCTGATCCAGACAGTGGCTGAGCTGCTCCCTCATTTCA 4800  
Db 4741 CTTCTCTGACAGCTGAGGCTGATCCAGACAGTGGCTGAGCTGCTCCCTCATTTCA 4800  
QY 4801 CCTTAGTGAACCTTAGATCTCTCTCCCAAGTAATGTTGACAGGCCAAATTTCAATACC 4860  
Db 4801 CCTTAGTGAACCTTAGATCTCTCTCCCAAGTAATGTTGACAGGCCAAATTTCAATACC 4860  
QY 4861 ATGTCAGATTTGAATGTATTTAAATGATGATTTAAGAGAACATGCTCTGTTCTGTT 4920  
Db 4861 ATGTCAGATTTGAATGTATTTAAATGATGATTTAAGAGAACATGCTCTGTTCTGTT 4920  
QY 4921 CCGTGTGGTTCCAGACAGTGGTTCTTGGTTGTTTCCCTGGCTAAGAGTCAAGTGGCC 4980  
Db 4921 CCGTGTGGTTCCAGACAGTGGTTCTTGGTTGTTTCCCTGGCTAAGAGTCAAGTGGCC 4980  
QY 4981 AAGATTAAGATTTATCTGGGGGAAAGAAATTTTAAAAATTTAAACTTAAGAT 5040  
Db 4981 AAGATTAAGATTTATCTGGGGGAAAGAAATTTTAAAAATTTAAACTTAAGAT 5040  
QY 5041 GTTTTAAGCTTAACCCGTAATTTGGGATGGAAGAGACAGACCCGCTGAACAGCTGT 5100  
Db 5041 GTTTTAAGCTTAACCCGTAATTTGGGATGGAAGAGACAGACCCGCTGAACAGCTGT 5100  
QY 5101 ATTTACAGACACACCCAGTGGTGTAAGACCAACAAATGACAGTCTGATGAAAGC 5160  
Db 5101 ATTTACAGACACACCCAGTGGTGTAAGACCAACAAATGACAGTCTGATGAAAGC 5160  
QY 5161 TCTAAGACCAATGTTGAAAGAGTCTCCAGTTACTGAACAGATGAAAGAGAGCTGTGAG 5220  
Db 5161 TCTAAGACCAATGTTGAAAGAGTCTCCAGTTACTGAACAGATGAAAGAGAGCTGTGAG 5220  
QY 5221 AGGCGTGTAAACATTAACAAATATTTTCCCTGTTTCTTTGTTAAAAACCAACATGG 5280  
Db 5221 AGGCGTGTAAACATTAACAAATATTTTCCCTGTTTCTTTGTTAAAAACCAACATGG 5280  
QY 5281 TTCACTGAATTCATGAATTTGAAAGAAATATTTCTAATTTAAATTAAGTCCCTTTAG 5340  
Db 5281 TTCACTGAATTCATGAATTTGAAAGAAATATTTCTAATTTAAATTAAGTCCCTTTAG 5340  
QY 5341 TTTGATCAGACAGCTTGAATCAGATCTCTTCTTCCCTGTCAGCTGACCTTCCCTTCC 5400  
Db 5341 TTTGATCAGACAGCTTGAATCAGATCTCTTCTTCCCTGTCAGCTGACCTTCCCTTCC 5400  
QY 5401 CCTCTCTCATTTCCCATACAGTCCCTATTTTCTCATCTTTTAAAAAATTAATTAAGTAC 5460  
Db 5401 CCTCTCTCATTTCCCATACAGTCCCTATTTTCTCATCTTTTAAAAAATTAATTAAGTAC 5460  
QY 5461 AGAACCAGGTAAGCCCTTATTTCTTAAATGTTTGGCAGCCACTTACCAATGCTAA 5520  
Db 5461 AGAACCAGGTAAGCCCTTATTTCTTAAATGTTTGGCAGCCACTTACCAATGCTAA 5520  
QY 5521 GTATTTGATTTTCAAGAAAAAATTTGATTTTACTGCGAAGAGAGCAAAAGTTAAGCC 5580  
Db 5521 GTATTTGATTTTCAAGAAAAAATTTGATTTTACTGCGAAGAGAGCAAAAGTTAAGCC 5580  
QY 5581 TTTGATACCAATCGAGCTTAAGATACCTGCTTTGGAACAGATTTATTTCTGTTCCCGCA 5640  
Db 5581 TTTGATACCAATCGAGCTTAAGATACCTGCTTTGGAACAGATTTATTTCTGTTCCCGCA 5640  
QY 5641 ACTCTGCTCCCAAAATGGGAGAAAGCCAGTGTGTTTAAATGATGACAGATATACGA 5700  
Db 5641 ACTCTGCTCCCAAAATGGGAGAAAGCCAGTGTGTTTAAATGATGACAGATATACGA 5700



Oy 5701 CAGATTAACTCTGCCATGCTTTTATTTTATTTTATTTTACAGCTGACTAAAGCCGA 5760  
 |||||||  
 Db 5701 CAGATTAACTCTGCCATGCTTTTATTTTATTTTATTTTATTTTACAGCTGACTAAAGCCGA 5760  
 Oy 5761 AGTTTGAAGTACATATAATCCCAATTATATGTAAACAGCAATATTTAGTTGAGA 5820  
 |||||||  
 Db 5761 AGTTTGAAGTACATATAATCCCAATTATATGTAAACAGCAATATTTAGTTGAGA 5820  
 Oy 5821 ACTTATGTTTTTATTTATTTATTTTGTGAGTATACATATTTGGAAATGACTCAA 5880  
 |||||||  
 Db 5821 ACTTATGTTTTTATTTATTTATTTTGTGAGTATACATATTTGGAAATGACTCAA 5880  
 Oy 5881 AATGAGTACTTCACTATTTAATTTAGATATCTTCAATACAGCTCTCCAAAGCTGTTT 5940  
 |||||||  
 Db 5881 AATGAGTACTTCACTATTTAATTTAGATATCTTCAATACAGCTCTCCAAAGCTGTTT 5940  
 Oy 5941 GTAAAGGATATCATGCTTGTATTAAGACCTAATTTGTAGACTTAAAGACTTTTATTTCT 6000  
 |||||||  
 Db 5941 GTAAAGGATATCATGCTTGTATTAAGACCTAATTTGTAGACTTAAAGACTTTTATTTCT 6000  
 Oy 6001 AAACCTGTGATTCGCTTATTAAGTATTTATCTATATCTATATGATGACAGCCGCTGA 6060  
 |||||||  
 Db 6001 AAACCTGTGATTCGCTTATTAAGTATTTATCTATATCTATATGATGACAGCCGCTGA 6060  
 Oy 6061 GGAACCAATTCCTGATTTTATTTATTTATTTCTTCTTATGAACCTTAGAAGACTA 6120  
 |||||||  
 Db 6061 GGAACCAATTCCTGATTTTATTTATTTATTTATTTCTTCTTATGAACCTTAGAAGACTA 6120  
 Oy 6121 CATGTTACTAAGCAGCCACTTTATGCTGTTT 6156  
 |||||||  
 Db 6121 CATGTTACTAAGCAGCCACTTTATGCTGTTT 6156

## RESULT 2

MMU39060 4878 bp mRNA linear ROD 25-APR-1997  
 LOCUS  
 DEFINITION Mus musculus glucocorticoid receptor interacting protein 1 (GRIP1)  
 mRNA, complete cds.  
 ACCESSION U39060  
 VERSION U39060  
 KEYWORDS U39060.1 GI:1853979  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
 AUTHORS Hong, H., Kohli, K., Trivedi, A., Johnson, D.L. and Stallcup, M.R.  
 TITLE 1 (bases 1 to 3560)  
 GRIPI, a novel mouse protein that serves as a transcriptional coactivator in yeast for the hormone binding domains of steroid receptors

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 93 (10), 4948-4952 (1996)  
 MEDLINE 96209838  
 REFERENCE 2 (bases 1 to 4878)

AUTHORS Hong, H., Kohli, K., Garabedian, M.J. and Stallcup, M.R.  
 TITLE GRIPI, a transcriptional coactivator for the AF-2 transactivation domain of steroid, thyroid, retinoid, and vitamin D receptors  
 JOURNAL Mol. Cell. Biol. 17 (5), 2735-2744 (1997)  
 MEDLINE 97265407  
 REFERENCE 3 (bases 1167 to 3560)

AUTHORS Stallcup, M.R.  
 TITLE Direct Submission  
 JOURNAL Submitted (20-OCT-1995) Department of Pathology, University of Southern California, 2011 Zonal Ave. HMR 301, Los Angeles, CA 90033, USA  
 REFERENCE 4 (bases 1 to 4878)

AUTHORS Hong, H. and Stallcup, M.R.  
 TITLE Direct Submission  
 JOURNAL Submitted (12-FEB-1997) Department of Pathology, University of Southern California, 2011 Zonal Ave. HMR 301, Los Angeles, CA 90033, USA

REMARK  
 COMMENT Sequence and feature updates by submitter  
 FEATURES On Feb 26, 1997 this sequence version replaced gi:1314284.  
 Location/Qualifiers

source  
 1. 4878  
 /organism="Mus musculus"  
 /strain="ICR outbred"  
 /db\_xref="taxon:10090"  
 /tissue\_type="brain"  
 /dev\_stage="newborn"  
 gene  
 1. 4878  
 /gene="Grip1"  
 204. 4592  
 /gene="Grip1"  
 /function="transcriptional co-activator for steroid receptors and nuclear receptors"  
 /note="hormone-dependent interaction with hormone binding domains of steroid receptors; transactivation activity in yeast and mammalian cells"  
 /codon\_start=1  
 /evidence=experimental  
 /product="glucocorticoid receptor interacting protein 1"  
 /protein\_id="AAC53151.1"  
 /db\_xref="GI:1853980"  
 /translation="MSGMGENTSPPSRAETRRKRECEPDGLSPKSTERNREDEK  
 YIEELADLIEFANFDIDFENFRPKCALIKETVQIHOIKOEKRAAANIDEVQSV  
 SSTQGVIDKDALGPMLEALDGFVFNLEGSVFPVSENVTOYLRIQNOELMKSVY  
 SILHVGHTFEFKMLIPKSMYNGSGMSGEPPRRSHFNCMLVPLPDSSEGHDSQ  
 EAHOKYEMOCFAVSOPRSIKEEGEDLOSLCYARRPKMERPLPSSESTRODL  
 QIKTSLDTSPIRAMKPGMEDIVRCLORHNOHEGESSYARRHHNHYLRGLARS  
 OIYFSLSDGTLYAAQTKSKILIRSDTINERPOLYSLHMLHREOVYCNMPDLGOAKG  
 KPLNPISSSPAHQALCSGNPGDMDTGSNINFPNNGKEDQMPKRFSGSGMNV  
 SGMOATTPQGSNVAALKMNSPQSSPGMNPQGVASSVLSPRQMSGTVGSPRI  
 PPSQPS  
 PASCILSHVGVCSSTGNSHYTNSSLNALQISGHVGLSSIASPDLKNGTNSP  
 VNMPPLPSKMGSLSDKDFGLYGEPSGEGTGAOEAOSCHPEQKGDSSMPQASGD  
 RAEGSHRLHDSKGGTKLLOLITTKSDMEPSPLPSISDPTKKSDTSGSLPGSGTHGS  
 LKEKHILHRLDSSSPVDLAKTAEATGKELSOBESSYAPGSEVYKOPAPAPKK  
 ENALRYLLEDDTDKIDGPEITTKLELDSKTDPASNTKILAKTVKEVSFEPSQ  
 PGSELNLELIDLONSQLEPDLPRPGAPTGSVKQALINDLMOTADSSPVA  
 GAOKAALRMSQSTFNNPBGQIGRLINONLIDITLOSPTGAGPFPTRNSPVTY  
 PPGMNGOGKLGSGNNGNNTGMSSTSRPMPGEMAPQAPAVRVCATTTGAM  
 NRPVGGIIRNPASITPMRANSGOGMOSLOMOSVLPSELENNMGSDPOLNOOAPR  
 NOTAPMPESTLIPDOASFSQNRPPSSPDILCPHPAESPDEBGLADOLTLAR  
 NFDLEIEDRLALGIPELYSQSAVDADQFSSQESSIMLEOKPVPFPQTSOAMAG  
 GYNPMQDNEHTMGQRNPNTTLRMQPRGLPTIIVONPQLOLQLOHROAOONNO  
 PLNMQISSVSNVNLJLRGVPTQAPINAMQAOQRELNOHROMOOQVOORTLM  
 MRGGLNVTSPMVAQGLPAMSNRIPQAAQOFPFPNPGISQDPDGTGATTPQ  
 SPLKSPRAHOSPMQOOSQANPAVOTPSDMNGAOSMGNSNFSQSPHPFQOAN  
 TSWSNMNMISVSMATNTGSLNSMONTQMSMTSVTSVPSGLPSMGPEDVNDPALR  
 GGNLFPQOLEMDMIKQBGDSKRYC"  
 BASE COUNT 1312 a 1381 c 1256 g 926 t 3 others  
 ORIGIN  
 Query Match 62.5%; Score 3846; DB 10; Length 4878;  
 Best Local Similarity 88.1%; Pred. No. 0;  
 Matches 4256; Conservative 0; Mismatches 553; Indels 20; Gaps 6;  
 Oy 1 GGGGCGCCAGCCTCGGTACAGCTTCGGCGCGAAGGTCAAGCGCGAGCGCGCA 60  
 |||||||  
 Db 44 GGGGCGCCAGCCTCGGTACAGCTTCGGCGCGAAGGTCAAGCGCGAGCGCGCA 103  
 Oy 61 CCGACGGCGTGACGAGCGAGCCGAGCGGATTTCTTTGATTTGGCTACACTTATAGATC 120  
 |||||||  
 Db 104 CCGACGGCGTGACGAGCGAGCCGAGCGGATTTCTTTGATTTGGCTACACTTATAGATC 161  
 Oy 121 TTCTGCATCTGTTACAGCACAGTTCGATATGTGTTCAAGATGATGGGATGGAGAA 180  
 |||||||  
 Db 162 TTCTGCATCTGTTACAGCACAGTTCGATATGTGTTCAAGATGATGGGATGGAGAA 221  
 Oy 181 AATACCTTGACCCCTCCAGGCGAGAGACAAGAAAGCCCAAGGATGCTCCGACCAATT 240  
 |||||||  
 Db 222 AATACCTTGACCCCTCCAGGCGAGAGACAAGAAAGCCCAAGGATGCTCCGACCAAGT 281  
 Oy 241 GGACCCAGCCCAAGGACACTGAAAGAGTAACTGTGAACAGGAAATATATATATA 300  
 |||||||  
 Db 282 GGACCCAGCCCAAGGACACTGAAAGAGTAACTGTGAACAGGAAATATATATATA 341

QY 301 GAAGAACTGCAGAGTTGATTTTGCAGAAATTTATGATATAGACAACTTAACTTCAA 360  
DB 342 GAGAGAGCTGGCCATCTGATCTTTCGCAAACTTTATGATATGACAACTTCAACTTCAA 401  
QY 361 CCGACAAATGTGCATCTTAAAGAAAGCTGTGAAGCAATGTGTCAGATCAAGAACAA 420  
DB 402 CTTGACAAATGTGCATCTTAAAGAAAGCTGTGAAGCAATGTGTCAGATCAAGAACAA 461  
QY 421 GAGAAAGCAGCAGCTGCCAATAGATGAAGTGCAGAGTGCAGATGTATCTCTACAGGG 480  
DB 462 GAGAAAGCAGCAGCTGCCAATAGATGAAGTGCAGAGTGCAGATGTATCTCTACAGGG 521  
QY 481 CAGAGGTCTCATGCAAGAGATGGCGCTGAGCTATGATGCTTGAAGCCCTTGATGGTTC 540  
DB 522 CAGAGGTCTCATGCAAGAGATGGCGCTGAGCTATGATGCTTGAAGCCCTTGATGGTTC 581  
QY 541 TTTCTTGTATGAAACCGGAGGCAAGCTGTGTTGTGTCAGAGATGTGACACATAT 600  
DB 582 TTTCTTGTGTGAACCTGGAGAGCGATGTGTGTTCTGTGACAGAAATGTGACACATAT 641  
QY 601 CTAGAGTATTAACCAAGAGAGCTGATGATGACAAAAGTGTATATAGCATCTTGATGGG 660  
DB 642 CTAGAGTATTAACCAAGAGAGCTGATGATGACAAAAGTGTATATAGCATCTTGATGGG 701  
QY 661 GACCACACGGAATTTGTCAAAAACCTGCTGCCAAAGTCTATAGTAAATGGGGATCTTGG 720  
DB 702 GACCACACGGAATTTGTCAAAAACCTGCTGCCAAAGTCTATAGTAAATGGAGATCTCTGG 761  
QY 721 TCTGGGGAACCTCGAGGGGGAACAGCATACCTTCAATGTGAGGATGGTAAACCT 780  
DB 762 TCTGGGGAACCTCGAGGGGGAACAGCATACCTTCAATGTGAGGATGGTAAACCT 821  
QY 781 TTACTGATTCAGAGAGAGGGGTGATGATTAACAGGAAGCTCATGAAATATGAAGT 840  
DB 822 TTGCGAGATTCAGAGAGAGGGGTGATGATTAACAGGAAGCTCATGAAATATGAAGT 881  
QY 841 ATGCAAGTCTGCTGCTGTCTCTCAACCAAGTCCATCAAGAGAGAGAGATTTGAG 900  
DB 882 ATGCAAGTCTGCTGCTGTCTCTCAACCAAGTCCATCAAGAGAGAGAGATTTGAG 941  
QY 901 TCTGCTGATTTGCGGAGGGAAGAGATGTCATGAGAAAGACCATGTTCTTCCCTCA 960  
DB 942 TCTGCTGATTTGCGGAGGGAAGAGATGTCATGAGAAAGACCATGTTCTTCCCTCA 1001  
QY 961 TCAGAAAGTTTACTACTGCCAGATCTTCAAGGCAAGATCAAGTCTGTGATTAACAG 1020  
DB 1002 TCAGAAAGTTTACTACTGCCAGATCTTCAAGGCAAGATCAAGTCTGTGATTAACAG 1061  
QY 1021 ACCATGAGAGCAGCATGAAGACAGGCTGGAGAGACCTGTTAAGAGGTATTCAGAG 1080  
DB 1062 ACCATGAGAGCAGCATGAAGACAGGCTGGAGAGATCTGTTAAGAGGTATTCAGAG 1121  
QY 1081 TTTCATGCGGAGATGAGAGAGATGTCGTATGCTTAAGGAGATCATCATGAGTAA 1140  
DB 1122 TTTCATGAGACAGATGAGAGAGAGTCTATCATATGCCAGAGGATCATCATGAGTAA 1181  
QY 1141 CTGAGAACAGAGATGGCATTCAGTCAATCTATGTTTTCTTGTCTGATGGACATCTT 1200  
DB 1182 CTGAGAACAGAGATGGCATTCAGTCAATCTATGTTTTCTTGTCTGATGGACATCTT 1241  
QY 1201 GTTGTCTGACAAAGCAAGAGCAAACTCATCGTTCTGACACTACTAATGAACCTCA 1260  
DB 1242 GTTGTCTGACAAAGCAAGAGCAAACTCATCGTTCTGACACTACTAATGAAGCTCA 1301  
QY 1261 GTAATATCTTTTACATATGCTTCAAGAGACAGAAATGTGTGATGATTCGGAGATCTG 1320  
DB 1302 GTAATATCTTTTACATATGCTTCAAGAGACAGAAATGTGTGATGATTCGGAGATCTG 1361  
QY 1321 ACTGAGCAAAAGATGGGGAAGCCAGTGAATCAATTAAGTCTTAACAGCCCTGCCATCAG 1380  
DB 1362 ACTGAGCAAAAGATGGGGAAGCCATTTGAATCAATTAAGTCTTAAGCAAGCCCTGCCAC 1421

QY 1381 GCCCTGTGAGTGGGAAGCCAGGTGACAGACATGACCTCTCATGACATATTAATTTTCCC 1440  
DB 1422 GCCCTGTGAGTGGGAAGCCAGGTGACAGACATGACCTCTCATGACATATTAATTTTCCC 1481  
QY 1441 ATAAATGGCCCAAGGAACAAATGGGGATGCCATGGGGAGGTTGGTGTCTGGGGGA 1500  
DB 1482 ATGAATGGCCCAAGGAACAAATGGGGATGCCATGGGGAGGTTGGTGTCTGGGGGA 1541  
QY 1501 ATGAACCATGTGTGACGATGACAGCAAGCAACCATCTTACAGGGTATTAATATGACACTCAA 1560  
DB 1542 ATGAACCATGTGTGACGATGACAGCAAGCAACCATCTTACAGGGTATTAATATGACACTCAA 1601  
QY 1561 ATGAACAGCCCTCTACAAAGCAGCCCTGGCATATGATCAGAGACGCCACTCTCATGCTT 1620  
DB 1602 ATGAACAGCCCTCTACAAAGCAGCCCTGGCATATGATCAGAGACGCCACTCTCATGCTT 1661  
QY 1621 TCACCAAGGATGATGATGACAGCCCTGGAGTGGGGAGCCCTGGAATCCACACAGTACG 1680  
DB 1662 TCACCAAGGATGATGATGACAGCCCTGGAGTGGGGAGCCCTGGAATCCACACAGTACG 1721  
QY 1681 TTTTCCCTGACAGGAAGCTTGCATTTCCCTGTGAGGATTTTGCAGACAGCAGGAAATAGC 1740  
DB 1722 TTTTCCCTGACAGGAAGCTTGCATTTCCCTGTGAGGATTTTGCAGACAGCAGGAAATAGC 1781  
QY 1741 CATAGTATATCAACAGACTCCCTCATGATGCACTTACAGCCCTCAGCGAGGGGCAGCGGGT 1800  
DB 1782 CATAGTATATCAACAGACTCCCTCATGATGCACTCAGCAGCCCTCAGCGAGGGGCAGTGGGTC 1841  
QY 1801 TCATTAAGGATCATGTTGGGCTTACACAGACTTAAATATGGGCAATTTGCAAACTCCCA 1860  
DB 1842 TCATTAAGGATCATGTTGGGCTTACACAGACTTAAATATGGGCAATTTGCAAACTCCCA 1901  
QY 1861 GTTAATATGAATCTCTCCCTCAGCAGCAAGATGGGAAGCTTGGACTCAAAAAGCTGTTT 1920  
DB 1902 GTTAATATGAATCTCTCCCTCAGCAGCAAGATGGGAAGCTTGGACTCAAAAAGCTGTTT 1961  
QY 1921 GGACTATATGGGAGCCCTCTGAAGGTACAACTGAGACAGCAGAGACAGTCCATCT 1980  
DB 1962 GGACTATATGGGAGCCCTCTGAAGGTACAACTGAGACAGCAGAGACAGTCCATCT 2021  
QY 1981 GGAAGAGCAAAAGGAACAAATGACCCCAACCTGCCCCGGGCGTGAGAGTGAAGAGCT 2040  
DB 2022 GGAAGAGCAAAAGGAGCCCAATGATTCAGCATGCCCCAGGCGGACAGGAGAGGCT 2081  
QY 2041 GAGGGGACAGAGACTGTGACAGCAAGAGGCGCAGACCAATCTCTGAGTGTGAC 2100  
DB 2082 GAGGGGACAGAGCGGGGTGATGACAGAAAGGCGCAGACCAATCTCTGAGTGTGAC 2141  
QY 2101 ACCAATCTGATATGATGAGAGCCCTGCGCTTGAAGCACTCTTGTGCGATTAACAACAA 2160  
DB 2142 ACCAATCTGATATGATGAGAGCCCTTCAACCTTCCAGAGTCTTGTGCGACACAAACAG 2201  
QY 2161 GACTCCACAGTATGATGAGAGCCCTGCGCTTGAAGCACTCTTGTGCGACACAAACAG 2220  
DB 2202 GACTCCACAGGAGGAGCTTGGCTGGGCTGGGTCCACCGATGACCTGCGTCAAGGAGAG 2261  
QY 2221 CATTAATTTTGCACAGACTCTTGCAGAGACAGAGTTCCTGTGACTGTGGCAAGTTA 2280  
DB 2262 CATTAATTTTGCACAGACTCTTGCAGAGACAGAGTTCCTGTGACTGTGGCAAGCTG 2321  
QY 2281 ACAAGCAGAGCCACAGGCAAAAGACCTGAGCAGAGAGTCCAGCAGCAGCTCTGATCA 2340  
DB 2322 ACAAGCAGAGCCACAGGCAAAAGACCTGAGCAGAGAGTCCAGCAGCAGCTCTGATCA 2381  
QY 2341 GAAGTGACTATTAACAAAGAGCGGGTGAAGCCCAAGAAAGAAAGATGATGACTTGGC 2400  
DB 2382 GAAGTGACTATTAACAAAGAGCGGGTGAAGCCCAAGAAAGAAAGATGATGACTTGGC 2441  
QY 2401 TATTTGCTAGATTAAGATGATTAAGATATTTGTTTACAGAAATTAACCCCAAACTT 2460  
DB 2442 TATTTGCTAGCAAAAGATGATTAAGATATTTGTTTACCGGAATTAACCCCAAACTT 2501  
QY 2461 GAGAGACTGACAGTAAAGACAGATCTCGCCAGTAAACAAATTAATAGCATGAAAAC 2520

Dp	2502	GAGGACGTGGACAGTATAGACAGATCTCGCAGTAAACAAATTAATGCTATGAAACT	2561
Qy	2521	GAGAGAGAGGAGATGAGCTTTGAGCCTGGTGACACAGCCGCGAGTACAGTACCTGGACAACTTG	2580
Dp	2562	GTTGAGGAGGAGGTGAGCTTTGAGGCCAGTGAACAGCCCTGGGAGGACTGAGCAACTTG	2621
Qy	2581	GAGGAGATTTTGGATGATTTGGCAGAAATAGTCAATTACACAGCTTTTCCAGACGAGG	2640
Dp	2622	GAAGAGATTTTGGATGATTTGGCAGAAACAGTACAGTTTCCACAGCTTTTCCAGACGAGG	2681
Qy	2641	CCAGGCCCTCTCGATACAGTTGACAGCAAGCCATCATATATAGACTCATGCAATC	2700
Dp	2682	CCAGAGACTCTCTCGATGAGTATGTTGACAGCAAGCCATCATATATAGACTCATGCAATC	2741
Qy	2701	ACAGCTGAAGAAAGCCCTGTCAACCTGTTTGGAGCCAGCAAAACAGCACTGCAATTTCA	2760
Dp	2742	ACAGCTGACAGCAAGTCCCGTCCACCTGCCGGAGCCGCAAGGACGCACTGCGATGCA	2801
Qy	2761	CAGAGCACTTTAATAAACCCACAGCAGGAGCACTGGGAGGTTATTTGCCAAACAGAAAT	2820
Dp	2802	CAGAGCACTTTAATAAACCCACAGCAGGAGCACTGGGAGGTTATTTGCCAAACAGAAAT	2861
Qy	2821	TTTACCATTTACATCACATTTGCCAAACCCCACTGGTGTGTGACCTTTTCCACATCAGA	2880
Dp	2862	TTTACCATTTACATCACATTTGCCAAACCCCACTGGTGTGTGACCTTTTCCACATCAGA	2921
Qy	2881	AACAGTATGCTCCATACATCAGTATACCTTCAGCCAGGAATATATGGATTAACAAGGATGATA	2940
Dp	2922	AACAGTATGCTCCATACATCAGTATACCTTCAGCCAGGAATATATGGATTAACAAGGATGATA	2961
Qy	2941	GGAAGCCAGGAATAATTTAGGGAACAGTATGACAGAGAAATGTTGTTACAGTGTCTTCTGG	3000
Dp	2982	GGAAGCCAGGAATAATTTAGGGAACAGTATGACAGAGAAATGTTGTTACAGTGTCTTCTGG	3041
Qy	3001	CTTACTATGCTCATCTGGAGAAATGGGACCGCAGAGTGTGGCTGTGATGATCACTGTGCT	3060
Dp	3042	CTTACTATGCTCATCTGGAGAAATGGGACCGCAGAGTGTGGCTGTGATGATCACTGTGCT	3101
Qy	3061	GCTTACACACAGTATGCAATGAACCGGCGCAGTCCAAAGAGATATATTCGGAACCCAGCAGCC	3120
Dp	3102	GCTTACACACAGTATGCAATGAACCGGCGCAGTCCAAAGAGATATATTCGGAACCCAGCAGCC	3161
Qy	3121	AGCATCCCATGAGGCCAGCAGCCAGCCTGGGCCAAAGACAGACGCTTCACTCAGGTC	3180
Dp	3162	AGCATCCCATGAGGCCAGCAGCCAGCCTGGGCCAAAGACAGATGCTTCACTCAGGTC	3221
Qy	3181	ATGAATATAGGCGCATCTGAATTAAGAGATGAACATGATGGGGGACCTCAGTATAGCCACAA	3240
Dp	3222	ATGAATATAGGCGCATCTGTGATTAAGAGATGAACATGATGGGGGACCTCAGTATATCAACAG	3281
Qy	3241	CAAGCTCTCTCAATCAGATGCGCCGATGGCCGGAAGCATCTCTGCTATATGACCGAGCG	3300
Dp	3282	CAGGCCCTCTCGAACAACACTGCCCGGTGCTGAGAGCATCTCTGCTATATGACCGAGCA	3341
Qy	3301	TCTTTTCCAGCAAAACAGCAGCCATTTGGCAGTTTCCAGATGATCTTGTATGTCCA	3360
Dp	3342	TCTTTTCCAGCAAAACAGCAGCCATTTGGCAGTTTGAATGAGGCTCTTGGGATATCCAGAACTGGTC	3401
Qy	3361	CATCTCTCAGCTATGATCTCCGAGTATGAGGAGCTCTCTGGACCAAGCTGATATCTGGCC	3420
Dp	3402	CATCTCTCAGCAAGATGCGCCAAACGATGAGGGGCGCTCTTCTTGCACACCTGATATCTGGCC	3461
Qy	3421	TTTGGGGAATTTTATGATGGCGTGGAGGAGATTTGATATAGAGCTTATGGGAATCCGGAATCTGTC	3480
Dp	3462	TTTGGGGAATTTTATGATGGCGTGGAGGAGATTTGATATAGAGCTTATGGGATATCCAGAACTGGTC	3521
Qy	3481	AGCCAGAGCCAGCAGTATAGTATCCAGAAAGTTCTCAAGTACAGATATCCAAATCATCTGCTG	3540
Dp	3522	AGCCAGAGCCAGCAGTATGATATGAGAGAGTTCTCAAGTACAGATATCCAGAACTATATGCTG	3561
Qy	3541	GAGCAGAGGCGCCGTTTCCACAGCAGTATGATCTCAGGCAAAATGGCCCGAGGCT	3600

Db	3582	GAACAGAACCCCCCTTTTCCACAGCAGTACGATCTCAGCCAAATGGCCCCAGGGT	3641
Qy	3601	AGCTATTCCCATGCGAAGATCCMAACTTTCACACCATGGGACAGCGGCTTAGTTATGCC	3660
Db	3642	GGGTAAATCCCATGCAAGATCCAAACTTTCACACCATGGGACAGCGGCGCAATTTACAC	3701
Qy	3661	ACACTCCGATGACGCCAGACCGGGCCTCAGGCCACCGGGCTAGTGCAGAACGACCA	3720
Db	3702	ACACTCCGATGACGCCAGCGGCGCTCAGGCCACAGGCGCTTTACAGAACCGACCA	3761
Qy	3721	AATCAACTGAAGCTTCACCTTCAGTACGATCGGCTCCAGGCACAGAGAAATGGCCAGCACTT	3780
Db	3762	AACCAACTGAGACTTCAGTTAGCTTACGACCGCCTCCAGGACAGAGAAATGGCCAGCACTT	3821
Qy	3781	ATGAATCAATAGCAGAAATGTTTCCATATGTGAACCTTACTCTGAGGCGCTGGATGCCACA	3840
Db	3822	ATGAATCAATAGCAGAGTGTTCCTTCATATGTGAACCTTACTCTGAGGCGCTGGATGCCACT	3881
Qy	3841	CAGGCACTTATTATGACACAGATGCTGGCCACAGACAGAGGGAATCTGTAACCAAGCAT	3900
Db	3882	CAGGCTCTATTATATGACAGATGCTGGCCACAGAGGCAAGGGAATCTCTCAACCAACAC	3941
Qy	3901	CTTCCGACAGACAAATGCATCAGCAACAGCAAGTTCCAGCAACGAATTTGATGATGACA	3960
Db	3942	CTTCCGACAGAGCGCAGAT-----GCAGCAGCAGGTGACGACGAGGCGCTGATGATGACA	3995
Qy	3961	GGACAGGGTTGAATTTGACACCAACGATGATGGCTGCTAGTGGTATGCGACCAATTATG	4020
Db	3996	GGACAGGGCTTGAATTTGACACCAACGATGATGGCTGCTAGTGGCTTACCAGCAGCATG	4055
Qy	4021	AGCAACCTCGGATTTCCAGGCAATGACAGCAGTTCCTCCAACTACAGCA	4080
Db	4056	AGCAATCCCCGGATCCCCAGGCAATGCGCACAGTATTCCTCCGAACCTACGCA	4115
Qy	4081	ATTAAGTCAGCAACCTGTATTCAGGCTTTACTGGGGGTACAGATCTCCCAAGCCACTATG	4140
Db	4116	ATTAAGTCAGCAACCTGTATTCCTGGCTTTACTGGGGGTACAGATCTCCCAAGCTCTTAATG	4175
Qy	4141	TCACCCCGAATGGCAGATACACAGATGCCATATGCAACAGTCTCAGGCCAACCCAGCC	4200
Db	4176	TCACCCCGAATGGCAGATACATATCAGAGTCCCATATGACAGAGTCTCAAGCCAAACAGCC	4235
Qy	4201	TATCAGGCCCTCCGACATTAATGAGTGGCGCAGGGGAACATGGGGCGGAAACAGCATG	4260
Db	4236	TACACACCCACCTCAGACATGAATGATGGGCACAGGGGAGCATGGGTGGAAACAGCATG	4295
Qy	4261	TTTTTCCAGCAGTCCCCACCACTTTGGGGCAGCAAGCAACACAGCATGTACAGTAAC	4320
Db	4296	TTCTCTCACACAGTCCCCACCACTTTGGGGCAACAGCAACACAGCATGTATTAATTAAC	4355
Qy	4321	AACATTAACATCAATGTGTCCATGGGGAGGCAACACAGTGGGCTGAGCAGCATGAACCAAG	4380
Db	4356	AACATTAACATCAATGTGTGTCCATGGGAGGCAACACAGTGGGCTGAGCAGCATGAACCAAG	4415
Qy	4381	ATGACAGGACAGATGACATGACCTCAGTACCTCCGTCTTAACGTACAGGGCTGTCTCC	4440
Db	4416	ATGACATGCGCAGATGAGCAGTGAACCTCAGTACCTCCGTCTTAACGTACAGGGCTGTCTCC	4475
Qy	4441	ATGGGTCCCGAGCAGGTTAATGATCTCTGTCTGAGGGGAGGCAACCTGTTCCTCAAAACAG	4500
Db	4476	ATGGGTCCCGAGCAGGTCTATGATACCTCTGTCTGAGGGGAGGCAACCTTTTCCAAACCA	4535
Qy	4501	CTGCCCTGGAATGGATTTGATTAAGCAGGAGGGGAGCAACAACAGGAATATTTCTCAGAC	4560
Db	4536	CTGCCCTGGAATGGACATGATCAAGCAGGAGGGAGATGCAATCTTCGGAAATATCTGTAACCC	4595
Qy	4561	TGCTGAAGCAGTTCCTTCAGCTGACCGGGCTCACTTGTCTCAAAACACTTCCAGTCT	4620
Db	4596	TGGAGAAACTGTCTGTGATCTTTCTTCACACCACTGGGGTTA--CAAACTTTACAGTCT	4653
Qy	4621	GGAGAGCTGTCTATTATTGTTTCAACCAACTGACTGCCAGCGGTTCTGCTAG-----	4675
Db	4654	GGAGAGCTGTCTCTTTCTTTGTGT--TCCCACTGTCACTGTCCGCGCAGTCTTCCACAGGACATA	4711

[illegible][illegible]



Db	3117	TCGCCATGGGAGCCAAAGCCACCTTGGCCAAAGACAGATGCTTCAGTCTCAGGTCAAGA	3176
Qy	3185	ATATAGGCCATCTGAATTAGAGATGACATGGGGGACCTCAGTATAGCCAAACAAG	3244
Db	3177	ACATAGGCCCTTCTCAGTTGTAGATATACATGTGGAGACCTCAGTATATCAACAGCAGG	3236
Qy	3245	CTCTCCAAATACAGACTGGCCCAATGGGCTGAAAGATCTCTGGCTATAGACCGAGGGCTTT	3304
Db	3227	CCCTCCGAACCAACATCTGCCCCGTGGCTGTAGAGCATCTCTGCTTATAGACAGGCATGCT	3296
Qy	3305	TTGGCAGCCAAACAGGACGCCATTTTGGCAGTTCTTCAGATGACATGCTGTATGTCCATTC	3364
Db	3297	TTGGCAGCCAGAACAGGACACCTCTGGCAGCTCCCTGATGACCTGGTGTCCATTC	3356
Qy	3385	CTGAGCTGAGTCTCCAGTATGAGGAGGCTCTCTGGACACGCTGATCTGGCTTGC	3424
Db	3357	CTGAGAGAGATGCCCAAGGAGTGGGGGCTCTTCTTACCAAGCTGATCTGGCTTGC	3416
Qy	3425	GGATTTTGTAGTGGCTGGAGAGATTGATAGACCTTTAGAAATACCCGAACTGTCAACC	3484
Db	3417	GGAACTTCGATGGCTTGAGGAGATTGATAGAGCTCTGGGGATACCAAGACTGGTCAACC	3476
Qy	3485	AGACCCAGCAGTATGATCCAGAACGTTCTCAATGTCAGATTCACATCATCTGTGAGC	3544
Db	3477	AGACCCAGAGCTGTGTGATGAGACAGCTTCTCAAGTCCAGAGTCCAGCATTAATGTGGAGC	3536
Qy	3545	AGAAGGGCCCGTTTCTCCAGCAGTATGTCATCTCAGCACAAATGGCCCGAGGTACT	3604
Db	3537	AGAAAGCCCCCGTTTCTCCACAGCAGTATGATCTCAGCACAAATGGCCCGAGGTGCT	3596
Qy	3605	ATTTCCCATATGCAAGATCCAAACTTTACACACATGGGAGGAGGGCTATGTATGCCACAC	3664
Db	3597	ATATATCCATATGCAAGATCCAAACTTTACACACATGGGAGGAGGGCCAAATTTACACACAC	3656
Qy	3665	TTCGATATGACGCCACAGACCGGGCTCAGGCGCCACGGGCTACTGAGAAACCAAGCCAAATC	3724
Db	3657	TTCGATATGACGCCACAGGCGGCTCAGGCGCCACAGGCAATTTGTGAGAAACCAAGCCAAAC	3716
Qy	3725	AACATAGACTTCAACTTCACAGCATCGGCTCCCAAGCAGCAGAGATGCGCAGCCACTTATGA	3784
Db	3717	AACATAGACTTCAACTTCACAGCATCGGCTCCCAAGCAGCAGAGATGCGCAGCGGCTTATGA	3776
Qy	3785	ATCAATATCAGCAATGTTTCCATATGTAACTTGACTGTAGGCGCTGAGTATCCACACAGG	3844
Db	3777	ATCAGATATCAGCAAGTTTCCAAATGTAACTTGAAGCTCTTGAGGCGCTGAGTATCCACAGG	3836
Qy	3845	CACATTTAATATGCACAGATCTGGGCCACAGACAGAGGGAATTCGTAACACGATCTTC	3904
Db	3837	CTCCTATTAATATGCACAGATCTGGGCCACAGAGGGAATTCGTAACACGATCTTC	3896
Qy	3905	GACAGAGCAATATCATCATCAGCAACAGCAAGTTGACACGAACCTTGTATGATGAGAGAC	3964
Db	3897	GCGAGAGCAAGATATCAGCAGAG-----GTGCGAGGAGGGGACTTTGTATGATGAGAGAGC	3950
Qy	3965	AAGGTTGAATATATCAGCAAGCATGATGGCTCTTGTGTATGTCAGAGCAATATATGAGCA	4024
Db	3951	AGGGCTGAATATATATCAGCAGCATGATGGCTCTTGTGTATGTCAGAGCAATATATGAGCA	4010
Qy	4025	ACGCTGGATTCGCCAGGCAAAATGCACAGCAGTTTGCATTTCTCCAAACTACGGAATTA	4084
Db	4011	ATCCCCGGATTCGCCAGGCAAAATGCACAGAGTTTCCATTTCTCCGAACTACGGAATTA	4070
Qy	4085	GTCAGCAACCTGATCCAGGCTTTATCTGGGGCTATACAGACTCCCAAGACCCACTTATGTAC	4144
Db	4071	GTCAGCAACCTGATCCAGGCTTTATCTGGGGCTATCAACTCCCAAGAGTCTCTTAATGTTC	4130
Qy	4145	CCCAATATGGACATATACAGAGTCCCATGTATGTACAAAGTCTCAGGCCAAACCCAGCTATC	4204
Db	4131	CCCGGATATGGACATATCTCAAGTCCCATATGTAGAGCAAGTCTCAAGCCAAACCCAGCTATC	4190
Qy	4205	AGGCCCCCTCGACATTAATATGATGGGCGAGGGGAACATGGGCGGAAACAGCATGTTTT	4264

Db	4191	AGCCAGCTTCAGACATGTAATGGATGGGCAACAGGGGAGCATGGGTGGAAACAGCATGTTCT	4250
Qy	4265	CCACAGCACTCCCCACACACACTTTTGGGCGACAAAGCAACACAGCATGTACGTAAACAA	4324
Db	4251	CACAGCACTCCCCACACACTTTTGGGCAACAAAGCAACACAGCATGTATGTATTAACAA	4310
Qy	4325	TGACATCAATATGTGTCATGGCGAACACACAGGGGCGATGAGCAGCATGAAACAGATGA	4384
Db	4311	TGAACATATAGTGTGCGATGGCAACCAACAGGGGGGCTTATGACGACATGAAACAGATGA	4370
Qy	4385	CAGGACAGATCAGCATGACCTCAGTACCTCCGTCTACGTCAAGGCGTGTCTCCATGG	4444
Db	4371	CAGGCAATATAGCATGACATGACCTCAGTACCTCCGTCTACGTCAAGGCGTGTCTCCATGG	4430
Qy	4445	GTCCCGACAGGCTTATGATGATCTGCTCTGAGGGGAGGCAACTGTGTTCCCAACACAGCTGC	4504
Db	4431	GTCCCGACAGGCTTATGATGATGATCTGCTGAGGGGAGGCAACTGTTTCCCAACCAACTGC	4490
Qy	4505	CTGGAATGGATATGATTATAGCAGGAGGAGACACAAACCGGAAATATTGCTGCACATGCT	4564
Db	4491	TTGGAATGGACATATATCAAGGAGGAGGAGATGATCTCGGAAATATCTGTGACCTGGA	4550
Qy	4565	GAAGCAGATTGCTCTTACGTGAACCGGGGCTCACGTGCTAAAACACTTCTGAGCTGGAG	4624
Db	4551	GAACCTGTCTGCATCTTTTTCAAACCCACTGGGCTT--ACAAACATTTCACGATGTGAG	4608
Qy	4625	AGCTGTGTCTATTTTGTTCAAACCCAACTGACCTGGCCAGCGGTTCTGTGCTAGA-----	4676
Db	4609	AGCTGCGTCTTTTGTGT--TGCCACCTGACATGCGCGCAAGTTCTCCAGACATATGACG	4666
Qy	4677	---GCACAGAGCGCTGGCGCTTGCTTCCAGGCTGGCGTCCACTGGGCTGTGGCAGAGGA	4733
Db	4667	CAGACAGTCGGGCGCTGGGCGCGAGCATPAAGGCTGTGGCTTGCGCCACCGGAGGA	4726
Qy	4734	GCTGCTCTCTCTTGAACAGTCTGAAGCTGCATCCAAAC	4773
Db	4727	GTGCTCTCTCCC--GACAGCCTGAGGCTGCCTCCAAAC	4764

```

RESULT 4
AF136943
LOCUS AF136943 4398 bp mRNA linear ROD 12-APR-1999
DEFINITION Rattus norvegicus transcriptional intermediary factor 2 (TIF2)
ACCESSION AF136943
VERSION AF136943
KEYWORDS rRNA, complete cds.
SOURCE AF136943.1 GI:4581052
Norway rat.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Cranialta; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 4398)
leers,J., Treuter,E. and Gustafsson,J.-A.
Mechanistic principles in NR box-dependent interaction between
nuclear hormone receptors and the coactivator TIF2
Mol. Cell. Biol. 18 (10), 6001-6013 (1998)
98414616
2 (bases 1 to 4398)
leers,J., Treuter,E. and Gustafsson,J.-A.
Direct Submission
Submitted (24-MAR-1999) Department of Biosciences, Karolinska
Institute, Haelsovaegen 7, Huddinge 14157, Sweden
Location/Qualifiers
1. 4398
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
1. 4398
/gene="TIF2"
1. 4398
/gene="TIF2"
/codon_start=1
/function="nuclear hormone receptor coactivator"

```









OY	4003	GGTATGCCAGCAACATATGACGAACCCTGGGATTTCCCCAGGCAAATGCAACAGCATTTTCCA	4062
Db	3841	GGTTTACAAGCAGCCATGAGCGAACCCCGGGATCCCGCAAAGCCAAAGCCAGCTTTTCCA	3900
OY	4063	TTTTCTCCAAACTATGAGGAATAAGTGAGCAACCTGATATCCAGGCTTAATGAGGGGCTACGACT	4122
Db	3901	TTTTCTTCCGAATACGGAATAGTAGTGACGAACCTGATCTCGGCTTTACTGAGACTTCAAACT	3960
OY	4123	CCCCAGAGCCCACTTATGTGCACCCCGGAATGGCACATACAGAAGTCCCATGATGCAAGAC	4182
Db	3961	CCCCAGAGTCCCTAATGTCTCCCGGAATGGCACATACGANAGTCCCATGATGCAAGAC	4020
OY	4183	TCTCAGGCCAACCCAGCCTATATGAGCCCCCTCCGACATPAATGATGAGGCCAGGGGAAAC	4242
Db	4021	TCTCAGGCCAACCCAGCCTATACCAAGCCCGCCTCAGACATTAAGATGGGACAGGGGAAGC	4080
OY	4243	ATGGCGGAAAACAGCATGTTTTTCCAGAGAGTCCCAACACACTTTTGGGAGCAACCAAC	4302
Db	4081	ATGGCGGAAAACAGCATGTTCTCACAAAGTCCCAACACACTTTGGGCAACAAACGAAAC	4140
OY	4303	ACCACAGATGTACAGTATACAAACATGTAACATATGTCATGCGAGCCAGCCAAACAGATGGC	4362
Db	4141	ACCACAGATGTACAATATACAAACATGAAATCAACATGTCCATATGGCAACCAACAGGCTGGC	4200
OY	4363	ATGACACAGATGTAACCAACATGACAGACAGATGACATGACCTCAGTGAACCTCCGTCTCT	4422
Db	4201	TTGACCAACATGTAAACCAATGTACAGGCCAGATGACAGATGACCTCAGTGAACCTCTGTGCT	4260
OY	4423	ACGTCAGGAGCTGTCTCATGGGTCCCGAGCAGGTAAATGATCCGTCTGTGAGGGAGAGC	4482
Db	4261	ACATAGAGGCTGTCTCATGGGTCTCTGACAGGTCAATGACCTTGCTCTGAGGGGAAGC	4320
OY	4483	AACCTGTT---CCCAACCCAGCTGCTGGAATGATGATTAATTAAGCAGAGGAGACACA	4539
Db	4321	AGCCTTTTCACCACAAACCAACTGCCTGGAATGACATGATCAACAGAGGAGATGGG	4380
OY	4540	ACACGGAAATTAATTTGGTGA	4557
Db	4381	TTCTGGAAATTAATCTGCTGA	4398
RESULT	5		
LOCUS	XLA243119	4871 bp	mRNA linear VRT 07-APR-2000
DEFINITION	Xenopus laevis mRNA for transcription intermediary factor 2 (tlif2 gene).		
ACCESSION	AJ243119		
VERSION	AJ243119.1	GI:5123920	
KEYWORDS	TFIIF gene; transcription intermediary factor 2.		
SOURCE	Xenopus clawed frog.		
ORGANISM	Xenopus laevis		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Piplidae; Xenopodinae; Xenopus.		
AUTHORS	1. (bases 1 to 4871)		
TITLE	de la Calle-Mustienes,E. and Gomez-Skarmeta,J.L. XRF2, a Xenopus homologue of the human transcription intermediary factor, is required for a nuclear receptor pathway that also interacts with CBP to suppress Brachyury and XMyoD		
JOURNAL	Mech. Dev. 91 (1-2), 119-129 (2000)		
MEDLINE	20171035		
RECORD	2 (bases 1 to 4871)		
AUTHORS	Gomez-Skarmeta,J.L.		
JOURNAL	Direct Submission		
FIELDS	Submitted (15-JUN-1999) Gomez-Skarmeta J.L., Biology, University of Chile, Las Palmeras 3425, Casilla 653, CHILE		
FEATURES	Location/Qualifiers		
source	1..4871		
gene	/organism='Xenopus laevis' /db_xref='taxon:8355' /dev_stage='10.5-11'		
	165..4715		

[illegible]

```

Db      549  CGTGAAGGAAATGTGGTGTGTTCTGAGAAATGTGACCCAAATATCTCAGATCAACACG 608
Oy      616  GAAGAGCTGTGAACAAAAGTGTATATAGCATCTGTGATGGGACCACAGCAATTT 675
Db      609  GAAGAACTGATGATACTAGTGTGTACAGCATTTTGCATGTGGAGATCTCACAGGATTC 668
Oy      676  GTCAAAAACCTGTGCTCCAAAGTGTATAGTAATAGGGGATCTTGTGTGGCAACTCG 735
Db      669  ATCAAAAACCTGTGCTCCAAAGTGTATAGTAATAGGGG-----TGCT 710
Oy      736  AGCGGAACAGCATACCTTCATATGTCCGATGTGCTAATAACCTTTACCTGATTCAG 795
Db      711  AGCGGAATAGCCATACCTTTAACTGTGTATCTGTGTGAGCAATGATGATGAGTGG 770
Oy      796  GAGAGAGGTGATATTAACAGCAAGCATCAGAAAATATGAATATGATGATCTTCTGCT 855
Db      771  GAGGAAAGGCAGCATGTGTGAGAAACATCAGAAATAGCAAAATGATGATGATGCTTGT 830
Oy      856  GTCTCTCAACCAAGTCCATCAAGAAAGAGAGATTTGCAAGTCTGCTGATTTGC 915
Db      831  GTATCCCAACCAAGTCCATCAAGAAAGAGATTTGCAAGTCTGCTGATTTGT 890
Oy      916  GTGGCAAGAAGATTTCCCATGAAGAAAGACAGTCTTCCCTCATCAGAAATTTACT 975
Db      891  GTGGCCAGAAAGATTTCCCATGAAGAAAGAGAGGCTGTGCTCCCATCTGTGAGCTTC 950
Oy      976  ACTGCGAGAGATCTCCAGAGCAATCATCAGTCTGTGATACAGCAACATGAGAGAGCC 1035
Db      951  ACCGCTGAGAGATCTTCAAGTAAATATACATCTCTAGACAAACAGATGATGAGGCCCT 1010
Oy      1036  ATGAACACAGGCTGGGAGAGACCTGTGAAGAGTGTATTTCAGAAATTCATGCGCAGCAT 1095
Db      1011  ATGAGACACAGAGATGGAGAGACATGTGCGGCTTGATTCAGAGATTTCTCCACGAT 1070
Oy      1096  GAAGGAACATCTGTGCTTGTCTAGAGAGCATCATCATCAATGATGATGAGACAGATTG 1155
Db      1071  GATGGAGAA--ATATCTTACAGCAGAGAAACATCACCAAGATTTCTCAGAGAGGCGAT 1127
Oy      1156  GCATTTAGTCAATCTATCTTTTCTTTCTGTGATGGACCTGTTGCTGTGACAAAG 1215
Db      1128  GCGACCAAGTCTTTCTACGCGCTTTCTCTGTGCGATGTGATGACATGTTGTGGCCAAACA 1187
Oy      1216  AAGAGCAAACTCATCCGTTCTCAGACTACTAATGAACCTCACTTGTATATATCTTACAT 1275
Db      1188  AAGAGAGGCTCATGCGTCTACAGACTAATATGAACCTCCGCTGTGATTCATCAT 1247
Oy      1276  ATGCTTACAGAGAGAGAAATGTGTGTATATCCGATCTGACTGACGACAAAGCATG 1335
Db      1248  GTCTTTCAGAGAGAGAAATGTTGTGATTAATCAGAGATTTGCTGGCAAGCATG 1307
Oy      1336  GGGAGAGCACTGAATCCAAATTAAGTCTAACAGCCCTGCCATAGGCGCTGTGACGAG 1395
Db      1308  GGAAGAGCATTAATTCAGTTCAGTCCAGCAGTCCAGCTCATCAGCAAAATGTACGAGTGG 1367
Oy      1396  AACCCAGGTGAGAGCATGACCTCTAGTAGCATATATAATTTTCCATAAATGGCCCAAG 1455
Db      1368  AACCCAGGTGAGATACGACCATTAAGATGAATATGAAATATGCAATAGTCCCAAG 1427
Oy      1456  GAACCAATGGGCAATGGCCATGGGCAAGTGTGTGCTGTGGGGAATGAACCATGTGTCA 1515
Db      1428  GAACCAATGGGCTTGGGATGTGAGGTGTTGTGTTCAAGGTGGAATGAACCATATCTCA 1487
Oy      1516  GGCATGCAACCAACCACTCTCAGAGTATGATGATGATGATGATGATGATGATGATGAT 1575
Db      1488  AGTTTGAACCAACCACTCTCAGAGTATGATGATGATGATGATGATGATGATGATGAT 1547
Oy      1576  CAAAGAGAGCCCTGGATGAATCCAGAGACAGCCCACTCCATGCTTTTCCACCAAGCATGCG 1635
Db      1548  CATGAGAGTCTGAGCATGGGAGGCTCAGCCTAATTTCTATCTTCCACCAAGCATGCT 1607
Oy      1636  ATGAGGCTGTGATGTGTGCAAGCCCTCGAATCCCAAGCATGATGATTTTCCCTGCGAGA 1695

```

---

```

Db      1608  GTCAGCCAGAGACTGGCAGAGAACTCCTCGATTTGCACCCGATGCTTTTCCCTGCAAGA 1667
Oy      1696  AGCTTGCAATTCCTCTGTGGAGTTTGCAGCAGCAGCAAGAAATAGCCATGATTAACCAAC 1755
Db      1668  AGTTTGATTCACCACTAGTGTGTGAGAGACAGCAAGAAATAGCCATGATTAACCAAC 1727
Oy      1756  AGCTTCCCTCAATGACTTTCAGGCTTCAGCAGAGAGGCGAGGGGTCTCATTTAGGGTATG 1815
Db      1728  AGCTTCTTAATGCACTTCAGGCTTCAGAGAGGGTCAAGG---CCCTTGGCCCCCA 1784
Oy      1816  TTGCTTACACAGACCTTAAATAATGGGCAATTTGCAAAAGCTCCCGATTAATTAATGCT 1875
Db      1785  TTGTCTCTCTCATATGTGAAAGGTGAATCTTCAGCATTCGCTGCAACATGAAACCT 1844
Oy      1876  CCCCACCTGACGAAGATGGGAGCTTGGACTCAAAAGACTGTTTGTGATATATGGGAG 1935
Db      1845  CCTCAGATTAAGAAAGATGGGAGCATAGACTTAAGAAATCTTTGTGTGTGAGAA 1904
Oy      1936  CCTCTGAAGGTCAACTGAGACAGAGAGAGCATGCCATCTGTGAGAGCAAAAGAA 1995
Db      1905  CAGCTGAAAGTCAACTGAGCAGAGGGAGTCCGCTGCTCAATGAGCAGAAAGAG 1964
Oy      1996  ACAATGACCCCAACCTGCCCCCGGTGAGCAGTGAAGAGAGCTGACGGGAGAGCAG 2055
Db      1965  TCGCGTGA-----AAACTGTCAAGTGTGTTGACAAAACAGAGGCCAAAGCAGA 2015
Oy      2056  CTGCATGACAGCAAGAGGCGAGACCAACTCTCAGCTGTGACCAACAAATCTGATCAG 2115
Db      2016  TTGCTTGATGGCAAAAGTGCAGAAAACCTGTGAACCTGTGACCAACAAATCTGACAG 2075
Oy      2116  ATGAGAGCTTGGCCCTTAGCCAG---CTCTTTGTGATTAACAAAGACTCCACAGT 2172
Db      2076  ATGAGAGCATCCAGCTACTAATACTTTGGGAGACTGAACAAAGACTCTTGTCA 2135
Oy      2173  AGCTTGC-----TGCTTGGGTACATGGAACCTCGGTCGAAGAGAGCAT 2223
Db      2136  AATTTTGATCAAAATGATGATGTGCTGCTCAGCAATGAACTCATGAAAGAGAGCAT 2195
Oy      2224  AAAATTTGACAGACTCTTTCAGAGACAGCATTCCTGCTGTGACTTGGCAAGTTAACA 2283
Db      2196  AAAATTTGACAGACTCTTTCAGAGACAGCATTCCTGCTGTGACTTGGCAAGTTAACA 2255
Oy      2284  GCAAGAGCCACAGGCAAAAGCTGTGAGAGAGTCCAGACGACAGCTCTGTGATGAA 2343
Db      2256  GCAGAGGCCACCGGAAAGAGTGTGAGAGAGTCCAAATGACAGAGCCCTGTTGAGAG 2315
Oy      2344  GTGACTATTAAACAAGCCGGTGAAGCCCAAGAAAGAGAGTACTACTTCCGTAT 2403
Db      2316  GTCACTATTAAAGCAGAGACAGTCAAGCCCAAGAAAGAGTACTACTTCCGTAT 2375
Oy      2404  TTGCTGATTAAGATGATATTAAGATATTGTTTACAGAAATTAACCCCAACTTGAG 2463
Db      2376  TTATTTGATTAAGATGATATTAAGATATTGTTTACAGAAATTAACCCCAACTTGAG 2432
Oy      2464  AGACTGACAGTAAAGACAGATCTGCCAGTAAACAAATTAATAGCAATGAAACTGAG 2523
Db      2433  CGAGCAGATTAACAAGGTGAGCCCTAGCAGCTGTCAAAATTAATGCGCTGTCAAAAGAG 2492
Oy      2524  AAGGAGAGATGAGCTTTGAGCTGTGAGACAGCTGTGAGCTGTGAGCTGTGAGCACTTGAG 2583
Db      2493  AAGGAGAGGCAAACTTTGAGACACTGACACAGCCGGGTCTATTTTGTATTAACCTGAT 2552
Oy      2584  GAGATTTTGGATGATTTGCAAGATAGTCAATTAACCAAGCTTTTCCAGACAGAGGCCA 2643
Db      2553  GAAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2612
Oy      2644  GCGGCGCTCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2703
Db      2613  G-----ATGCAACTCAGGAGTAAACAAGCCATGATGATGATGATGATGATGATGATGAT 2666
Oy      2704  GGTGAAAACAGCCCTGTGACACCTGTTGGAGGCC---AGAAAACAGAGCTGCGAATTTTCA 2760
Db      2667  GTGAAAACAGCAAGTCTCCCTGCGTGGGCCCAAAAGCAGAGATGCTCCGGATGACAG 2726

```





FEATURES	source
NOTE: This is a 'working draft' sequence. It currently consists of 13 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.	
1	987: contig of 987 bp in length
988	1087: gap of 100 bp
1088	18618: contig of 17531 bp in length
18619	18718: gap of 100 bp
18719	29940: contig of 6222 bp in length
24941	25040: gap of 100 bp
25041	30091: contig of 5051 bp in length
30092	30191: gap of 100 bp
30192	50359: contig of 20168 bp in length
50360	50459: gap of 100 bp
50460	75475: contig of 25016 bp in length
75476	75575: gap of 100 bp
75576	86946: contig of 11371 bp in length
86947	87046: gap of 100 bp
87047	104572: contig of 17526 bp in length
104573	104672: gap of 100 bp
104673	109999: contig of 5327 bp in length
110000	110099: gap of 100 bp
110100	125482: contig of 15383 bp in length
125483	125582: gap of 100 bp
125583	129768: contig of 4186 bp in length
129769	129868: gap of 100 bp
129869	173364: contig of 43496 bp in length
173365	173464: gap of 100 bp
173465	181403: contig of 7939 bp in length.
location/Qualifiers	
1.	181403
/organism="Homo sapiens"	
/db_xref="taxon:9606"	
/chromosome="8"	
/map="8"	
/clone="RP11-152C15"	
/clone_lib="RPC1-11 Human Male BAC"	
BASE COUNT	49970 a 40664 c 38806 g 50648 t 1315 others
ORIGIN	
Query Match	25.9%: Score 1592.8; DB 2; Length 181403;
Best Local Similarity	99.5%: Pred. No. 0;
Matches 1608: Conservative	0: Mismatches 7; Indels 1; Gaps 1;
QY 4542	ACGGAATATTTGCGACACTGCTGAAGCAGTGTCTTCAGCTGACCGGCTCCTTG 4601
Db 144116	ACAGAAATATTTGCGACACTGCTGAAGCAGTGTCTTCAGCTGACCGGCTCCTTG 144057
QY 4602	CTCAAAACACTTCCAGTCTGAGAGAGCTGTGCTATTTTTCACCAACCACTGACCTGCCA 4661
Db 144056	CTCAAAACACTTCCAGTCTGAGAGAGCTGTGCTATTTTTCACCAACCACTGACCTGCCA 143997
QY 4662	GCCGGTCTGCTAAGCAGAGAGAGCGCTGGCCCTGGTCCAGGGGGGCTCCACTCGGCT 4721
Db 143996	GCCGGTCTGCTAAGCAGAGAGAGCGCTGGCCCTGGTCCAGGGGGGCTCCACTCGGCT 143937
QY 4722	GTCGACAGAGAGAGCGCTCTTCTTCTTACAGTCTGAGAGCTCGCATCCAGACAGTCTGC 4781
Db 143936	GTCGACAGAGAGAGCGCTCTTCTTCTTACAGTCTGAGAGCTCGCATCCAGACAGTCTGC 143877
QY 4782	AGTCTGTTCCCTGATTCACCTAGTGTGACCACTAGATCTCTCCCAAGTAATGTTG 4841
Db 143876	AGTCTGTTCCCTGATTCACCTAGTGTGACCACTAGATCTCTCCCAAGTAATGTTG 143817
QY 4842	ACAGGCAATTTCCATTCACATTCAGATTTGAATGATTTAAATGTATTTAAAGAGA 4901
Db 143816	ACAGGCAATTTCCATTCACATTCAGATTTGAATGATTTAAATGTATTTAAAGAGA 143757

QY 4902 ACCATGCTCTTGTCTGTTCCCTGTTCCAGACACTGGTTTCTTGCTTTGTTCC 4961  
 |||||||  
 Db 143756 ACCATGCTCTGTTGTTCTGTTCCCTGTTCCAGACACTGGTTTCTTGCTTTGTTCC 143697  
 |||||||  
 QY 4962 TGGCTAACAGCTAGTGCACCAAGATTAAAGTTTATCTGGGGGAAAGAAAGATTTT 5021  
 |||||||  
 Db 143696 TGGCTAACAGCTAGTGCACCAAGATTAAAGTTTATCTGGGGGAAAGAAAGATTTT 143637  
 |||||||  
 QY 5022 AAAAAATTAACTAAAGATGTTTAAAGCTTAAGCTGAAGCTGAATTTGGATGGAAGAGACAG 5081  
 |||||||  
 Db 143636 AAAAAATTAACTAAAGATGTTTAAAGCTTAAGCTGAAGCTGAATTTGGATGGAAGAGACAG 143577  
 |||||||  
 QY 5082 ACACCGTGCAGACGCTGTATTTTACAGACACACCCAGTGGGTGAAGAACCAACAGCTCAC 5141  
 |||||||  
 Db 143576 ACACCGTGCAGACGCTGTATTTTACAGACACACCCAGTGGGTGAAGAACCAACAGCTCAC 143517  
 |||||||  
 QY 5142 AGTCGATCTCTGTAAGAGCTCTTAAGACCATGTTGGAAAGTCTCCAGTTACTGAACAG 5201  
 |||||||  
 Db 143516 AGTCGATCTCTGTAAGAGCTCTTAAGACCATGTTGGAAAGTCTCCAGTTACTGAACAG 143457  
 |||||||  
 QY 5202 ATGAAGAAGACCTGTGAGAGGCGTGTAACTATAGCAATATTTTCTGTTTTC 5261  
 |||||||  
 Db 143456 ATGAAGAAGACCTGTGAGAGGCGTGTAACTATAGCAATATTTTCTGTTTTC 143397  
 |||||||  
 QY 5262 TTTGTAAAAACAACCTGTTCACTGAATCATGAATTGAGAAATATTTTCAATTC 5321  
 |||||||  
 Db 143396 TTTGTAAAAACAACCTGTTCACTGAATCATGAATTGAGAAATATTTTCAATTC 143337  
 |||||||  
 QY 5322 TAAATTAAGTCCCTTTTATGTTTATGATCAGACAGCTTGAAATGACATCTCTTCCCTGTC 5381  
 |||||||  
 Db 143336 TAAATTAAGTCCCTTTTATGTTTATGATCAGACAGCTTGAAATGACATCTCTTCCCTGTC 143277  
 |||||||  
 QY 5382 AGCTGACCTCTCTCCCTCTCTCATATCCCATACCTCTATTTTCTTCTTTC 5441  
 |||||||  
 Db 143276 AGCTGACCTCTCTCCCTCTCTCATATCCCATACCTCTATTTTCTTCTTTC 143217  
 |||||||  
 QY 5442 AAAAAATATATAGCTTACAGAAACAGTAAAGCTTTATTTCTTAAATGTTTTGCCA 5501  
 |||||||  
 Db 143216 AAAAAATATATAGCTTACAGAAACAGTAAAGCTTTATTTCTTAAATGTTTTGCCA 143157  
 |||||||  
 QY 5502 GCCACTTACCAATTTGCTAGTATTTGAAATTTTCAAAAAAATGCAATTTACTGCGCAGG 5561  
 |||||||  
 Db 143156 GCCACTTACCAATTTGCTAGTATTTGAAATTTTCAAAAAAATGCAATTTACTGCGCAGG 143097  
 |||||||  
 QY 5562 AGAAGAGCAAAAGTTAAGGCTTGTATACCAATCGAGCTAAGATACCTGCTTTGGAAGCATG 5621  
 |||||||  
 Db 143096 AGAAGAGCAAAAGTTAAGGCTTGTATACCAATCGAGCTAAGATACCTGCTTTGGAAGCATG 143037  
 |||||||  
 QY 5622 TTTATTTCTGTTCCCGACCAACTCTGGCTCCAAATGCGAG -AAACGCCAGTGTGTTAA 5680  
 |||||||  
 Db 143036 TTTATTTCTGTTCCCGACCAACTCTGGCTCCAAATGCGAG -AAACGCCAGTGTGTTAA 142977  
 |||||||  
 QY 5681 ATTGATAGCAGATATCAGCAGACATTTAACTCGCCATGTTTATTTGTTGTTTT 5740  
 |||||||  
 Db 142976 ATTGATAGCAGATATCAGCAGACATTTAACTCGCCATGTTTATTTGTTGTTTT 142917  
 |||||||  
 QY 5741 AGCAGTCTGACTAAGCCGAGTTTGTAGGTACATAAATCCAAATTTATATGAACA 5800  
 |||||||  
 Db 142916 AGCAGTCTGACTAAGCCGAGTTTGTAGGTACATAAATCCAAATTTATATGAACA 142857  
 |||||||  
 QY 5801 AGCAATATTTAAGTGAAGACTTATGTTTAAATTTGTAATTTTGTGAGATATACA 5860  
 |||||||  
 Db 142856 AGCAATATTTAAGTGAAGACTTATGTTTAAATTTGTAATTTTGTGAGATATACA 142797  
 |||||||  
 QY 5861 TATGTGGAATTCACCAAAATAGGTAATCTAGATTAATTAATGATATCTCATRAGA 5920  
 |||||||  
 Db 142796 TATGTGGAATTCACCAAAATAGGTAATCTAGATTAATTAATGATATCTCATRAGA 142737  
 |||||||  
 QY 5921 ATGTCTCTAAAGGTGTTTGTAAAGATATCATGCTGATTTAGACCAATTTGTAGA 5980  
 |||||||  
 Db 142736 ATGTCTCTAAAGGTGTTTGTAAAGATATCATGCTGATTTAGACCAATTTGTAGA 142677  
 |||||||

QY 5981 CTTAAGACTTTTATTTTCTAAACCTTGATCTCTGCTTAAAGTCAATTAATCA 6040  
 |||||||  
 Db 142676 CTTAAGACTTTTATTTTCTAAACCTTGATCTCTGCTTAAAGTCAATTAATCA 142617  
 |||||||  
 QY 6041 TATGATATGACAGCCGCTGTAAGAACCAATTTGATTTTATATGTTATATTTCTTCT 6100  
 |||||||  
 Db 142616 TATGATATGACAGCCGCTGTAAGAACCAATTTGATTTTATATGTTATATTTCTTCT 142557  
 |||||||  
 QY 6101 AATGAACCTTGAAGACATACATGTTACTAAGCAGCCACTTTATGTTGTTTT 6156  
 |||||||  
 Db 142556 AATGAACCTTGAAGACATACATGTTACTAAGCAGCCACTTTATGTTGTTTT 142501  
 |||||||

RESULT 8  
 AF254984 71701 bp DNA linear HTG 24-OCT-2001  
 LOCUS Homo sapiens chromosome 8 clone RP4-611F18, WORKING DRAFT SEQUENCE,  
 DEFINITION 20 unordered pieces.  
 ACCESSION AF254984  
 VERSION AF254984.4 GI:16356859  
 KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE  
 AUTHORS Men.G., Baumgart,C., Blechschmidt,K., Dete,M., Jahn,N.,  
 1 (bases 1 to 71701)  
 Men.G., Baumgart,C., Blechschmidt,K., Dete,M., Jahn,N.,  
 Lehmann,R., Menzel,U., Polley,A., Reichwald,K., Schilhabel,M.B.,  
 Schudy,A., Siddiqui,R., Taudien,S., Rosenthal,A. and Platzter,M.  
 Chromosome 8 genomic sequence  
 Unpublished  
 2 (bases 1 to 71701)  
 Polley,A., Men.G., Baumgart,C., Dete,M., Jahn,N., Schilhabel,M.,  
 Menzel,U. and Rosenthal,A.  
 Direct Submission  
 Submitted (12-Apr-2000) Genome Analysis, Institute of Molecular  
 Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany  
 On Oct 24, 2001 this sequence version replaced gi:14327871.

COMMENT  
 TITLE JOURNAL  
 AUTHORS  
 REFERENCE  
 ORGANISM

Center: Institute of Molecular Biotechnology  
 Center code: IMB  
 Web site: http://genome.imb.jena.de/  
 Contact: gscj-submit@genome.imb.jena.de  
 Project Information  
 Center project name: H266  
 Center clone name: RP4-611F18

----- Summary Statistics -----  
 Sequencing vector: M13; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 50971 bases at least Q40  
 Consensus quality: 57915 bases at least Q30  
 Consensus quality: 63233 bases at least Q20  
 Quality coverage: 3.29 x in Q20 bases; sum-of-contrigs

----- Sequence Quality Assessment -----  
 This entry has been annotated with sequence quality  
 estimates computed by the Phrap assembly program.  
 All manually edited bases have been reduced to quality 10.  
 Quality levels above 40 are expected to have less than  
 1 error in 10,000 bp.  
 Base-by-base quality values are not generally visible from the  
 Genbank flat file format but are available as part  
 of this entry's ASN.1 file.

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 20 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will





QY 5921 ATGTCCTCAAGCTGTTTGTAAAGATATCAATGCTTGAATAGACCTAATTTGTAGA 5980  
 |||||||  
 DB 67558 ATGTCCTCAAGCTGTTTGTAAAGATATCAATGCTTGAATAGACCTAATTTGTAGA 67617  
 |||||||  
 QY 5981 CTTAAGACCTTTATTTTCTAAACCTTGATCTGCTTAAAGTCAATTAATCACTCA 6040  
 |||||||  
 DB 67618 CTTAAGACCTTTATTTTCTAAACCTTGATCTGCTTAAAGTCAATTAATCACTCA 67677  
 |||||||  
 QY 6041 TATGATATGACGCGCTGTAGAACCAATCTGATTTTATATGTTATATCTTCTCT 6100  
 |||||||  
 DB 67678 TATGATATGACGCGCTGTAGAACCAATCTGATTTTATATGTTATATCTTCTCT 67737  
 |||||||  
 QY 6101 AATGAACCTTGAAGACCTACATGTTACTACTAGACGCGCATTGTATGCTTTT 6156  
 |||||||  
 DB 67738 AATGAACCTTGAAGACCTACATGTTACTACTAGACGCGCATTGTATGCTTTT 67793  
 |||||||

RESULT 9  
 AC090280 AC090280 86554 bp DNA linear HTG 23-MAY-2001  
 LOCUS Homo sapiens chromosome 8 clone RP4-611F18 map 8, WORKING DRAFT  
 DEFINITION  
 AC090280  
 AC090280  
 AC090280.2 GI:13958483  
 HTG: HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
 SOURCE  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo;  
 1 (bases 1 to 86554)  
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
 Homo sapiens chromosome 8, clone RP4-611F18  
 Unpublished  
 2 (bases 1 to 86554)  
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,  
 Barmann,N., Bastien,V., Boguski,M., Bouckhalter,B., Brown,A.,  
 Camarata,J., Campione,A., Choquet,Y., Colangelo,M., Collins,S.,  
 Collumore,A., Cooke,P., DeArnell,K., Dewar,K., Diaz,J.S.,  
 Dodge,S., Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J.,  
 Gargana,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,  
 Hags,B., Heathford,A., Horton,L., Hulme,W., Iller,I., Johnson,R.,  
 Jones,C., Karas,A., LaRoque,K., Lamazares,R., Landers,T.,  
 Lechoczy,J., Levine,R., Liu,G., Maclean,C., Macdonald,P.,  
 Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,  
 McPheters,R., Meldrum,J., Meneses,L., Mihova,T., Mlenka,V.,  
 Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,  
 O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,  
 Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,  
 Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,  
 Roy,A., Santos,R., Schauer,S., Schnupack,R., Seaman,S., Severy,P.,  
 Sounez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
 Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,  
 Travers,M., Travis,N., Trigglio,J., Vassiliev,H., Viel,R., Vo,A.,  
 Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zaitoun,J.,  
 Zembek,L., Zimmer,A. and Zody,M.

TITLE  
 JOURNAL  
 COMMENT  
 Direct Submission  
 Submitted (17-FEB-2001) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On May 6, 2001 this sequence version replaced gi:12957916.  
 All repeats were identified using RepeatMasker:  
 Smit,A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: MIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 Project Information  
 Center project name: L12502  
 Center clone name: 611.F.18  
 Summary Statistics  
 Sequencing vector: Plasmid; n/a; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731  
 Consensus quality: 83006 bases at least Q40  
 Consensus quality: 84782 bases at least Q30  
 Consensus quality: 85389 bases at least Q20  
 Insert size: 85000; agarose-fp  
 Insert size: 85854; sum-of-coverage  
 Quality coverage: 6.7 in Q20 bases; agarose-fp  
 Quality coverage: 6.6 in Q20 bases; sum-of-coverage  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 8 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \* 1  
 \* 3967 4066: gap of 100 bp  
 \* 4067 8136: contig of 4070 bp in length  
 \* 8137 8236: gap of 100 bp  
 \* 8237 12124: contig of 3888 bp in length  
 \* 12125 12224: gap of 100 bp  
 \* 12225 18749: contig of 6525 bp in length  
 \* 18750 18849: gap of 100 bp  
 \* 18850 31371: contig of 12522 bp in length  
 \* 31372 31471: gap of 100 bp  
 \* 31472 43391: contig of 11920 bp in length  
 \* 43392 43491: gap of 100 bp  
 \* 43492 71800: contig of 28309 bp in length  
 \* 71801 71900: gap of 100 bp  
 \* 71901 86554: contig of 14654 bp in length.  
 Location/Qualifiers  
 1. 86554  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="8"  
 /map="8"  
 /clone="RP4-611F18"  
 /clone\_lib="RP4 Human PAC library 4"  
 1. 3966  
 /note="assembly-fragment"  
 4067. 8136  
 /note="assembly-fragment"  
 8237. 12124  
 /note="assembly-fragment"  
 12225. 18749  
 /note="assembly-fragment"  
 18850. 31371  
 /note="assembly-fragment"  
 31472. 43391  
 /note="assembly-fragment"  
 43492. 71800  
 /note="assembly-fragment"  
 71901. 86554  
 /note="assembly-fragment"  
 /note="assembly-fragment"  
 clone\_end: 77  
 vector\_side: right"

BASE COUNT 25686 a 17956 c 17557 g 24632 t 723 others  
 ORIGIN

Query Match 20.7% Score 1274.2; DB 2; Length 86554;  
 Best Local Similarity 99.8% Pred. No. 0;  
 Matches 1276; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 1278 GCTTACAGAGACAGCAATGTGTGTGATGATCCGATCTGACATGACAAACGATGGG 1337  
 |||||||  
 DB 21477 GTATATCAGAGACAGCAATGTGTGTGATGATCCGATCTGACATGACAAACGATGGG 21536  
 |||||||  
 QY 1338 GAAGCCACATGATTCATTAAGCTCTTACAGACCTGCGCATACAGCCCTGTGACAGTGGAA 1397  
 |||||||  
 DB 21537 GAAGCCACATGATTCATTAAGCTCTTACAGACCTGCGCATACAGCCCTGTGACAGTGGAA 21596  
 |||||||



```

Oy 1398 CCCAGGTGAGACATGACCTCTAGTAGCAATATAATTTTCCCATAAATGGCCAAAGGA 1457
      |||
Db 21597 CCCAGGTGAGACATGACCTCTAGTAGCAATATAATTTTCCCATAAATGGCCAAAGGA 21656
Oy 1458 ACAAAATGGGCAATGCCATGGGAGTTTGTGGTTCTGGGGGAATGAACCATGTGTCAAG 1517
      |||
Db 21657 ACAAAATGGGCAATGCCATGGGAGTTTGTGGTTCTGGGGGAATGAACCATGTGTCAAG 21716
Oy 1518 CATGCAACCAACCACTCTCAGAGGTAGTAACTATGCACTCAAAATGAACAGCCCTCACA 1577
      |||
Db 21717 CATGCAACCAACCACTCTCAGAGGTAGTAACTATGCACTCAAAATGAACAGCCCTCACA 21776
Oy 1578 AAGCAGCCCTGGCATGAATTCAGAGACAGCCCACTCCATGCTTTACCAAGGATCGCAT 1637
      |||
Db 21777 AAGCAGCCCTGGCATGAATTCAGAGACAGCCCACTCCATGCTTTACCAAGGATCGCAT 21836
Oy 1638 GAGCCCTGGAGTGGCTGGAGCCCTCGAATCCCAACAGTCAGTTTCCCTGCAGGAAG 1697
      |||
Db 21837 GAGCCCTGGAGTGGCTGGAGCCCTCGAATCCCAACAGTCAGTTTCCCTGCAGGAAG 21896
Oy 1698 CTTGCAATCCCTCTGGGAGTTTTCAGAGACAGCAAGAAATAGCCATAGTTATACCAAG 1757
      |||
Db 21897 CTTGCAATCCCTCTGGGAGTTTTCAGAGACAGCAAGAAATAGCCATAGTTATACCAAG 21956
Oy 1758 CTCCCTCAATGCACTTTCAGAGCCCTCAGAGGAGGAGGCTCTCATTAGGCTCATCGTT 1817
      |||
Db 21957 CTCCCTCAATGCACTTTCAGAGCCCTCAGAGGAGGAGGCTCTCATTAGGCTCATCGTT 22016
Oy 1818 GGCCTTACACAGACCTAAATATGGCAATTTGCAAAATCTCCCAAGTTAATATGAATCTCTC 1877
      |||
Db 22017 GGCCTTACACAGACCTAAATATGGCAATTTGCAAAATCTCCCAAGTTAATATGAATCTCTC 22076
Oy 1878 CCCACTCAGCAAGATGGGAGCTTGAGCTCAAAAGACTGTTTGGACTATATGAGGGAGACC 1937
      |||
Db 22077 CCCACTCAGCAAGATGGGAGCTTGAGCTCAAAAGACTGTTTGGACTATATGAGGGAGACC 22136
Oy 1938 CTCTGAAGGTACAACTGGAGCAAGAGAGAGAGTCCATCCTGAGAGACAAAGGAAG 1997
      |||
Db 22137 CTCTGAAGGTACAACTGGAGCAAGAGAGAGAGTCCATCCTGAGAGACAAAGGAAG 22196
Oy 1998 AAATGACCCCAACTGCCCCCGCCGCTGAGAGAGTGAAGAGTGAAGAGAGAGACT 2057
      |||
Db 22197 AAATGACCCCAACTGCCCCCGCCGCTGAGAGAGTGAAGAGTGAAGAGAGAGACT 22256
Oy 2058 GCATGACAGCAAGAGGAGCAACAACTCTGAGAGTGTGACCAACCAATCTGATCAAT 2117
      |||
Db 22257 GCATGACAGCAAGAGGAGCAACAACTCTGAGAGTGTGACCAACCAATCTGATCAAT 22316
Oy 2118 GGAGCCCTCGCCCTTACGACCTCTTGTGCGATACAAACAAGACTCCACAGTAGCTT 2177
      |||
Db 22317 GGAGCCCTCGCCCTTACGACCTCTTGTGCGATACAAACAAGACTCCACAGTAGCTT 22376
Oy 2178 GCGTGTGTTCTGGTCTACATGGAACCTGCTCAAGAGAGCAATAAATTTTGCACAG 2237
      |||
Db 22377 GCGTGTGTTCTGGTCTACATGGAACCTGCTCAAGAGAGCAATAAATTTTGCACAG 22436
Oy 2238 ACTCTGAGAGACAGCAAGTCCCTGCTGAGACTTGGCCAAAGTTAACAGCAAGCCACAGG 2297
      |||
Db 22437 ACTCTGAGAGACAGCAAGTCCCTGCTGAGACTTGGCCAAAGTTAACAGCAAGCCACAGG 22496
Oy 2298 CAAAGACCTGAGCAGAGATCCAGAGACAGCTCTGATCAGAAAGTCAATTAACA 2357
      |||
Db 22497 CAAAGACCTGAGCAGAGATCCAGAGACAGCTCTGATCAGAAAGTCAATTAACA 22556
Oy 2358 AGAGCCGGTGAGCCCAAGAAAGAGAGATGCACTACTGCTATTTTGTAGATTAAGA 2417
      |||
Db 22557 AGAGCCGGTGAGCCCAAGAAAGAGAGATGCACTACTGCTATTTTGTAGATTAAGA 22616
Oy 2418 TGATACTTAAGATATTTGTTTACCAAGAAATTAACCCCAACTTGAGAGACTGAGACAGTA 2477
      |||
Db 22617 TGATACTTAAGATATTTGTTTACCAAGAAATTAACCCCAACTTGAGAGACTGAGACAGTA 22676

```

```

Oy 2478 GACAGATCTCGCCAGTAAACACAAATTAATAGCAATGAAGAACTGAGAGAGAGATGAG 2537
      |||
Db 22677 GACAGATCTCGCCAGTAAACACAAATTAATAGCAATGAAGAACTGAGAGAGAGATGAG 22736
Oy 2538 CTTTGAGCCTGGGTGACCAAG 2556
      |||
Db 22737 CTTTGAGCCTGGGTGACCAAG 22755

RESULT 10
AF207954/c 122924 bp DNA linear HTG 07-JUN-2001
LOCUS Homo sapiens chromosome 8 clone RP1-92A23 map 8q21, WORKING DRAFT
DEFINITION SEQUENCE, 8 unordered pieces.
ACCESSION AF207954
VERSION AF207954.3 GI:14327855
KEYWORDS HTG; PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 122924)
AUTHORS Reichwald,K., Baumgart,C., Blechschmidt,K., Dette,M., Jahn,N.,
Lehmann,R., Menzel,U., Polley,A., Schillhabel,M.B., Schudy,A.,
Siddiqui,R., Taudien,S., Wen,G., Siebert,R., Schlegelberger,B.,
Rosenthal,A. and Platzer,M.
Chromosome 8 genomic sequence
Unpublished
2 (bases 1 to 122924)
REFERENCE Reichwald,K., Wen,G., Schillhabel,M., Menzel,U., Baumgart,C.,
Dette,M., Jahn,N. and Rosenthal,A.
Direct Submission
Submitted (21-NOV-1999) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
On Jun 7, 2001 this sequence version replaced gi:18152106.
-----
Genome Center
Center: Institute of Molecular Biotechnology
Center code: IMB
Web site: http://genome.imb-jena.de/
Contact: gscj-submit@genome.imb-jena.de
-----
Project Information
Center project name: H265
Center clone name: RP1-92A23
-----
Summary Statistics
Sequencing vector: M13, 100% of reads
Chemistry: Dye-terminator Big Dye, 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 112126 bases at least Q40
Consensus quality: 116396 bases at least Q30
Consensus quality: 119383 bases at least Q20
Quality coverage: 5.23 x in Q20 bases; sum-of-contigs
-----
Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality 10.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
Genbank flat file format but are available as part
of this entry's ASN.1 file.
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 8065: contig of 8065 bp in length
* 8066 8165: gap of unknown length
* 8166 46558: contig of 38393 bp in length

```



Hagos, B., Hearford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Karatas, A., Laroque, K., Lamazares, R., Landers, F., Lehoczy, J., Levine, R., Liu, G., Maclean, C., Macdonald, P., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPheters, R., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunhkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S., Severy, P., Sougne, C., Spencer, B., Strange-Thomann, N., Stojanovic, N., Straus, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Travers, M., Travis, N., Triggillo, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission  
Submitted (09-MAR-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 124630)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Brown, A., Camarata, J., Campotiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Hearford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Laroque, K., Lamazares, R., Landers, F., Lehoczy, J., Levine, R., Liu, G., Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPheters, R., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunhkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S., Severy, P., Spencer, B., Strange-Thomann, N., Stojanovic, N., Straus, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Triggillo, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission  
Submitted (08-OCT-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
4 (bases 1 to 124630)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Bouknight, B., Brown, A., Camarata, J., Campotiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Hearford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Laroque, K., Lamazares, R., Landers, F., Lehoczy, J., Levine, R., Liu, G., Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPheters, R., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunhkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S., Severy, P., Spencer, B., Strange-Thomann, N., Stojanovic, N., Straus, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Triggillo, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission  
Submitted (12-DEC-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Dec 12, 2001 this sequence version replaced gi:15983548.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: MIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L12493  
Center clone name: 92\_A\_23  
-----  
Location/Qualifiers  
1. 124630  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="8"  
/map="8"  
/clone="RP1-92A23"  
/clone\_id="RP1 Human PAC library 1"  
159. .202  
/rpt\_family="AT\_rich"  
259. .307  
/rpt\_family="L2"  
279. .284  
/note="<30 qual SINGL region"  
932. .936  
/note="<30 qual SINGL region"  
1841. 1867  
/rpt\_family="(TTCA)n"  
complement(1878. .2040)  
/rpt\_family="Alusg"  
complement(2041. .2347)  
/rpt\_family="Alusg"  
complement(2349. .2657)  
/rpt\_family="Aluuo"  
3227. .3249  
/rpt\_family="AT\_rich"  
complement(3429. .3717)  
/rpt\_family="Alusg"  
3855. 3874  
/rpt\_family="(TGC)n"  
complement(3877. .4023)  
/rpt\_family="Alusg/x"  
4525. .4558  
/rpt\_family="AT\_rich"  
5389. 5476  
/rpt\_family="MIR3"  
complement(5958. .6261)  
/rpt\_family="Aluuo"  
complement(8011. .8311)  
/rpt\_family="Alusx"  
9089. 9118  
/rpt\_family="AT\_rich"  
complement(11149. .11426)  
/rpt\_family="Aluuo"  
11512. .11782  
/rpt\_family="L1M48"  
12189. .12501  
/rpt\_family="MERS8B"  
12617. .12623  
/note="<30 qual SINGL region"  
15270. .15338  
/rpt\_family="Aluud/FLAM"  
15341. .15505  
/rpt\_family="Alusg/x"  
15724. .15927  
/rpt\_family="MER2"  
15947. .15967  
/rpt\_family="(TA)n"  
17377. .17426  
/rpt\_family="(GA)n"  
18914. .18989  
/rpt\_family="MIR"  
complement(19636. .19938)

```
repeat_region /rpt_family="AluSx"
                complement(21513..21758)
repeat_region /rpt_family="MIR"
                2194..22017
repeat_region /rpt_family="AT-rich"
                complement(22060..22186)
repeat_region /rpt_family="MIR"
                complement(22543..22843)
repeat_region /rpt_family="AluSc"
                complement(22941..23128)
repeat_region /rpt_family="MER3"
                complement(23688..23729)
repeat_region /rpt_family="L1PA7"
                complement(24557..24637)
repeat_region /rpt_family="MIR3"
                complement(25039..25344)
repeat_region /rpt_family="AluSx"
                26351..26490
repeat_region /rpt_family="L1MA2"
                26615..26648
repeat_region /rpt_family="(CAT)n"
                complement(27973..28134)
repeat_region /rpt_family="MIR"
                complement(28392..28983)
repeat_region /rpt_family="AluJo"
                32586..32622
repeat_region /rpt_family="AT-rich"
                32738..32805
repeat_region /rpt_family="MIR"
                36137..36181
repeat_region /rpt_family="GA-rich"
                complement(36667..36800)
repeat_region /rpt_family="AluJo"
                complement(37421..37594)
repeat_region
```

Query Match 20.7% Score 1274.2: DB 9: Length 124630:  
Best Local Similarity 99.8% Pred. No. 0:  
Matches 1276; Conservative 0; Mismatches 3; Indels 0; Gaps 0:

```
QY 1278 GCTTCACAGAGAGAGAGATGTGTGTGATGATCCGATCTGACTGACAAAGATGGG 1337
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 13973 GTATATACAGAGACAGATGTGTGTGATGATCCGATCTGACTGACAAAGATGGG 13914

QY 1338 GAAGCCACTGATTCATTAATGCTCTTAACAGCCCTGCCATCAGGCCCTGTGACATGGAA 1397
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 13913 GAAGCCACTGATTCATTAATGCTCTTAACAGCCCTGCCATCAGGCCCTGTGACATGGAA 13854

QY 1398 CCCAGGTCAGACATGACCCCTCAGTGAATATATATTTTCCATTAATGGCCAAAGGA 1457
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 13853 CCCAGGTCAGACATGACCCCTCAGTGAATATATATTTTCCATTAATGGCCAAAGGA 13794

QY 1458 ACAATATGGGATGCCCCATGGGAGGTTTGGTCTGGGGAAATGAACCATGTGCAG 1517
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 13793 ACAATATGGGATGCCCCATGGGAGGTTTGGTCTGGGGAAATGAACCATGTGCAG 13734

QY 1518 CATGCAAGACACCATCTCTCAGGGTACATATATGCACTCAAAATGAACGCCCTCACA 1577
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 13733 CATGCAAGACACCATCTCTCAGGGTACATATATGCACTCAAAATGAACGCCCTCACA 13674

QY 1578 AAGCAGCCCTGGCATGAATCCAGAGAGCCACCTCCATGCTTCAACCAAGGATGGCAT 1637
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 13673 AAGCAGCCCTGGCATGAATCCAGAGAGCCACCTCCATGCTTCAACCAAGGATGGCAT 13614

QY 1638 GAGCCCTGAGTGGTGGAGAGCCCTCGAATCCACCCAGTCACTTTTCCCTCAGAGAG 1697
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 13613 GAGCCCTGAGTGGTGGAGAGCCCTCGAATCCACCCAGTCACTTTTCCCTCAGAGAG 13554

QY 1698 CTTCGATTCCTCTGTGGAGATTGGACAGACACAGAAATAGCCATAGTTATACCAACAG 1757
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 13553 CTTCGATTCCTCTGTGGAGATTGGACAGACACAGAAATAGCCATAGTTATACCAACAG 13494

QY 1758 CTCCTCATATGACCTTCAGGCCCTCAGCAGAGGGGCAAGGGGCTCATTTAGGTCATCGTT 1817
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
DB 13493 CTCCTCATATGACCTTCAGGCCCTCAGCAGAGGGGCAAGGGGCTCATTTAGGTCATCGTT 13434

QY 1818 GGGTTCACAGACACTTAAATATGGGCAATTTGCATAAATCCGCCAGTTAATATGATCTCC 1877
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 13433 GGGTTCACAGACACTTAAATATGGGCAATTTGCATAAATCCGCCAGTTAATATGATCTCC 13374

QY 1878 CCCACTCAGCAGATGGGAAGCTTGGACTCAAAAGACTGTTTGGACTATATGGGAGCC 1937
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 13373 CCCACTCAGCAGATGGGAAGCTTGGACTCAAAAGACTGTTTGGACTATATGGGAGCC 13314

QY 1938 CTCCTGAGGTACAACTGGAGCAACAGAGACACTCCATCCCTGGAGACCAAAAGGAAC 1997
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 13313 CTCCTGAGGTACAACTGGAGCAACAGAGACACTCCATCCCTGGAGACCAAAAGGAAC 13254

QY 1998 AATGAGCCCAACCTGCCCCCGCGCTGAGCAGTGAAGAGCTGAGGGCAGAGCAGACT 2057
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 13253 AATGAGCCCAACCTGCCCCCGCGCTGAGCAGTGAAGAGCTGAGGGCAGAGCAGACT 13194

QY 2058 GCATGACAGCAAAAGGCGAGACCAAACTCCTGACGTGCTGACCACCAATCTGATCAGAT 2117
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 13193 GCATGACAGCAAAAGGCGAGACCAAACTCCTGACGTGCTGACCACCAATCTGATCAGAT 13134

QY 2118 GAGGCCCTCGCCCTTGAAGCCAGCTCTTTGTCGATTAACAAACAAAGACTCCACAGTACT 2177
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 13133 GAGGCCCTCGCCCTTGAAGCCAGCTCTTTGTCGATTAACAAACAAAGACTCCACAGTACT 13074

QY 2178 GCGTGGTTCGGGCTTACACATGGAACCTCGCTCAAGAGAAAGCATAAATTTTGCACAG 2237
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 13073 GCGTGGTTCGGGCTTACACATGGAACCTCGCTCAAGAGAAAGCATAAATTTTGCACAG 13014

QY 2238 ACTCTTGACAGACAGCAGTTCCCTGTGACCTTGGCAAGTTAAACAGCAGAAAGCCACAG 2297
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 13013 ACTCTTGACAGACAGCAGTTCCCTGTGACCTTGGCAAGTTAAACAGCAGAAAGCCACAG 12954

QY 2298 CAAGAGCTGAGGCGAGAGTCCAGCAGACAGCTCCTGATCAGAGATGATATTAAACA 2357
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 12953 CAAGAGCTGAGGCGAGAGTCCAGCAGACAGCTCCTGATCAGAGATGATATTAAACA 12894

QY 2358 AGAGCCGCTGAGGCCCAAGAGAAAGATGACACTCTCTGATTTCTGTGATTAAGA 2417
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 12893 AGAGCCGCTGAGGCCCAAGAGAAAGATGACACTCTCTGATTTCTGTGATTAAGA 12834

QY 2418 TGATACTAAAGATATTGTTTACACGAATTAACCCCAACTTGAAGACTGACAGCTAA 2477
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 12833 TGATACTAAAGATATTGTTTACACGAATTAACCCCAACTTGAAGACTGACAGCTAA 12774

QY 2478 GACAGATCCTGCCAGTAAACAAATTAATAGCAATGAAGAGAGAGAGATGAG 2537
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 12773 GACAGATCCTGCCAGTAAACAAATTAATAGCAATGAAGAGAGAGATGAG 12714

QY 2538 CTTTGAAGCTGTGACAG 2556
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 12713 CTTTGAAGCTGTGACAG 12695
```

RESULT 12  
AF323989 4958 bp mRNA linear VRT 21-FEB-2001  
LOCUS Dario rerio transcriptional intermediary factor 2 mRNA, complete  
DEFINITION cds  
ACCESSION AF323989  
VERSION AF323989.1 GI:13022009  
KEYWORDS  
SOURCE zebrafish.  
ORGANISM Dario rerio  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;  
Cypriniformes; Cyprinidae; Dario.  
REFERENCE 1 (bases 1 to 4958)  
AUTHORS Tan,J.H., Kara,A.U. and Chan,W.K.  
TITLE Molecular cloning and characterization of zebrafish nuclear  
receptor coactivator TIF2  
JOURNAL Unpublished

REFERENCE 2 (bases 1 to 4958)  
AUTHORS Tan, J. H., Kara, A. U. and Chan, W. K.  
TITLE Direct Submission  
JOURNAL Submitted (25-NOV-2000) Biological Sciences, National University of  
Singapore, Lower Kent Ridge Road, Singapore 119260, Singapore  
FEATURES  
source Location/Qualifiers  
1. .4958  
/organism="Danio rerio"  
/db\_xref="taxon:7955"  
196..4713  
/note="nuclear receptor coactivator; TIF2/GRIPI"  
/codon\_start=1  
/product="transcriptional intermediary factor 2"  
/protein\_id="AK11608.1"  
/db\_xref="GI:13022010"

## CDS

/translation="MSAVGERSSDPARSEAKREKREKESDILGSPKSTKRNHEHS  
KYLEBELLEIFANNDIDNENKVPDKALIKETVKKIQRKIDEDKAAANDEYOKAD  
VSTGOSVIDKLDKPMALDDEFFVNMENIVSENVTOYLRYNOBELMNTAV  
YSILHVDHAEFIKILNPKSHVNGVPMWSENPENSHFTNCRLMLNPHSEAEEDHE  
AOKYETMOCFAVESEPKSKEEGEDFOSCLICVARYPMKREPMPLPTQVETTRDQD  
GKITSLDTSLRASMKPMGMDIVRRCIORPHLONDQISPAKRHOOCVIRHGAESP  
YRESLSDCTIVSATKSKLYRSTSTNEPOLYMSLHIIQRBPVPCGMSADLGNATMK  
PMPMSPNTAGSCTPQGDATISSNTSTPSPGAKREPMHREFCPTGMSHATM  
QAAIPQSGYPLKLSFGSPGPMSPRHRASDPAVAGSPRLPPQFSPASLHSPAM  
CTSTGNGGAGANHYGTSSSLNALQALSCBCHGSHGSLGSPDKLGSAAHSAVS  
HHLINKSVSEFGEONOSBOGTGODEGIDLPDEKNGIGOFENLGSDDPSHLDN  
KSHTKILQLLTTRKTEPIESTSPPMAGCEPCCKDGKNGSGGNGSGSHATSLRK  
KHLHRLQNSSSPVDLAKLTAETKELCQDSAGTAYPELAKOEVPSPKRDNA  
LNRLLDKDNVLKGIKMEPEIKKEEGKMTGIKETKDGIDREPESELDID  
DLNOSQGLFIDSRPLSPSAVDKQSIINILQMTGSGAMSPQOORALQTAASQON  
FPARPQGPAPRPPRSVSLDMNMPPKGAGQGVPMNNNNPYSIMOOGMMGNHNAV  
NOPMNAVAGVGNNGPRVMOQDQMGAGTGVSAAPATAPAGOSHLPQALHSRV  
PNPATGPMRPNSSQGRPMLOPOMMANOSNNAPYSPATAPAGOSHLPQALHSRV  
VMEIERYPENGSSMTGSGOEDVLCPPASBEPDDECALLSOLTSALKDDEETDR  
ALGIPALVGAQPEVDQPEQDPSMMMDKPPATGQOYASSPANNQAGYTPMDA  
TFHGLQOMQORPXYPMLRMQARPLRGPGVVPQNPTRLQLOHRLQAOONRPMAN  
QMGVSNMNLPLRPVNPQGTINAKMLAORRELSSHRLRQLODQOQOQOOR  
SLAMRAGCLTMRPMAAAGMGAMSNPRLPONLQOFPYPRYGTGSLSPSTSP  
FSPGPMPLPAPOLISHSAMGSLANQCMGLSLQCPAVPMQMGHAFOPSSG  
MSQSDGREGGATPOSLMSPRKHGHOSTMPOMOAGCNSSFOSSPDINGMPQGININ  
SMFTQOSPPQPSOANNMTNGMNLNVMAANSSNMGMGNGMNSMSMNSGSGSM  
ANMGPEQKCYC"  
247..453  
/note="Region: basic helix-loop-helix DNA-binding domain"  
568..747  
/note="Region: PAS A domain"  
1000..1185  
/note="Region: PAS B domain"  
2083..2097  
/note="Region: NR Box"  
2251..2265  
/note="Region: NR Box"  
2419..2433  
/note="Region: NR Box"  
4123..4139  
/note="Region: HAT-like domain"  
BASE COUNT 1284 a 1405 c 1343 g 926 t  
ORIGIN

Query Match 20.1%; Score 1238.4; DB 5; Length 4958;  
Best Local Similarity 60.8%; Pred. No. 0;  
Matches 2515; Conservative 0; Mismatches 1411; Indels 210; Gaps 22;

146 GCTGATATGTGTTCAAGATGAGTGGATGAGAAAAATACCTTGACCCCTCAGAGGCG 205  
179 GGTGATATGATCTGAGATGAGTGGTGGGAAAACTCTGTCGACCCCTCGACGCTCAG 238  
206 AGACAAGAAAGCGCAAGAGATGTC---CTGACCAACTTGAGACCCCAAGGAGACA 262  
239 AGGCTCAGAAAGCGCAAGAGAGGACCTCTGGGCTCAAGAAAGGAACTA 298  
263 CTGAAAAAGCTAATCTGTAACAGAAAAATAATATATAGAAAGACTTGAGTTGATTT 322

Db 299 CAGAAAAAGCTAATCTGTAACAGAAAAATAATATATAGAAAGACTTGAGTTGATTT 358  
Qy 323 TTGCAATTTTAAATGATATAGACAACTTAACTTCAAACTGACAAATGTGCATTTAA 382  
Db 359 TCGCAACTTCAATGATACATTTGCAACCTTCAACGCTGACAAATGTGCATTTAA 418  
Qy 383 AAGAACTGTGTAAGCAATCTGTCAGATCAACAAAGAAAGAAAGAGACCTGCAACA 442  
Db 419 AGGAAACCGTAAACAGATCTGGCAATTTAAAGACAAAGAAAGAAAGCTGTCGACAAATG 478  
Qy 443 TAGATGAAGTGCAGAAAGTCAAGATGATCTCTACAGGCGAGGCTGTCACAGATG 502  
Db 479 AGGATGAAGTGCAGAAAGTCAAGATGATCTCTACAGGCGAGGCTGTCACAGATG 538  
Qy 503 CGCTGGGCGCTATGATGCTTGAAGCGCTTGAATGGTCTTCTTGTAGTAACTGGAAG 562  
Db 539 CCTTGGACCAATGATGCTGGAAGCGCTTGAATGGTCTTCTTGTAGTAACTGGAAG 598  
Qy 563 GCAACGTTGTGTTGTGTCAGAAATGACAGATGACAGATATCTAAGTATACCAAGAAAGC 622  
Db 599 GCACATTGTGTTGTATCGGAAATGTGACGACGATACCTGCGCTATATACAGAAAGC 658  
Qy 623 TGATGAACAAAGTGTATATAGCATCTTGCATGTTGGGACACACAGGAATTTGTCAAA 682  
Db 659 TGATGAACACAGTGTGTCAGACATACGATGTAGAGACCATGCGAATTCATCAAAA 718  
Qy 683 ACCTGCTGCCAAAGTCTAATGATTAATGAGGAGATCTTGTGTCGCAACCTCGAGCGGA 742  
Db 719 ACCTGCTGCCAAAGTCTAATGATTAATGAGGAGATCTTGTGTCGCAACCTCGAGCGGA 778  
Qy 743 ACAGCATACCTTCATGTTGCGGATGCTGTAAGAAACCTTACCGATTCACAAAGAGAG 802  
Db 779 ACAGCATACCTTCATGTTGCGGATGCTGTAAGAAACCTTACCGATTCACAAAGAGAG 837  
Qy 803 GTCATATACCAAGCAAGTCAATGATTAATGAAATGAAATGCAATGCTGCTGCTC 862  
Db 838 -----CAGATCTCAGAGGACGACAGCAAGAAATGAGACTATGACAGTCTTGTGCTGTG 892  
Qy 863 AACCAAGTCAATCAAGAAAGAGAGAGAAATGTCAGTCTGCTGTAATTTGCGTGCA 922  
Db 893 AACCAAGTCAATCAAGAAAGAGAGAGAAATGTCAGTCTGCTGTAATTTGCGTGCA 952  
Qy 923 GAAGAGTCCCATGAAAGAAAGACAGTCTTCCCTCATCAAGAAAGTTCATCTGCTGCC 982  
Db 953 GCGAGTCCCATGAAAGAAAGACAGTCTTCCCTCATCAAGAAAGTTCATCTGCTGCC 1012  
Qy 983 AGATCTCCAGAGCAAGATCAAGTCTGATACAGACAGCAGCAGCAGCAGCAGCAGC 1042  
Db 1013 AGGACTCCAGAGCAAGATCAAGTCTGATACAGACAGCAGCAGCAGCAGCAGCAGC 1072  
Qy 1043 CAGGCTGGAGACCTGCTGAAGAGTGTATTCAGAAAGTTCATGCGCAGCATGAAGAG 1102  
Db 1073 CAGGCTGGAGACCTGCTGAAGAGTGTATTCAGAAAGTTCATGCGCAGCATGAAGAG 1132  
Qy 1103 AATCTGTCTCTAATGCTAAGAGCATCATGTAAGTACTGACAGCAAGATTTGCAATTTCA 1162  
Db 1133 A--CATTTCTTCCGCAAAAGACATCAGCAAGAGTGTATGAAGCAGCGGAGCGTTTCA 1189  
Qy 1163 GTCAATCTATGTTTCTCTTGTCTGATGACACTCTGTTGTCGACAAAGAGAGCA 1222  
Db 1190 GTCCATCTACCGGTTCTCTGTCGACAGGAGCAGCAGCAGCAGCAGCAGCAGCAG 1249  
Qy 1223 AACTATCCGTTCTCAGACTCAATATGAATCACTCAATCTTGAATATCTTATATATGCTTC 1282  
Db 1250 AACTATCCGTTCTCAGACTCAATATGAATCACTCAATCTTGAATATCTTATATATGCTTC 1309  
Qy 1283 ACAGAGCAAGATGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1339  
Db 1310 AGAGGAGCAAGATGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1369  
Qy 1340 AGCCATGATATCTAATGCTTGAACAGCCTGCGCATCAGAGCCTGTCAGTGGGAAC 1399

Db 1370 AACCCATGAACCCCATGAGCTCTCAAAACACAGC-----GGGTTCCTGCTGACCTCCCC 1423  
Qy 1400 CAGGTGAGACATGACCTCTAGTACGAATTAATTTTCCCATTAATGAGCCAAAGAAC 1459  
Db 1424 AAGGGAGAGAGCGCACCATTCAGACACACATCCACTTTCCCTCCCTGGTGGCCCA 1483  
Qy 1460 AAATGGGCAATGCCCATGGGAGTTTGTGTGTTCTGGGGGAAATGAACCATGTCTAGGCA 1519  
Db 1484 AGGAGCTGGTGCATGACACCGTTTGGCTGCCGGGAACCATGAGCCATTCTGGCAACCA 1543  
Qy 1520 TGCACCAACACACTCTCAAGGGTAGTAACTATGCACTCAAAAATGACAGCCCTCCACAA 1579  
Db 1544 TGCACACGGCCACCTCTCAAGGCAAGGGATCTCACTGAAGCTGACAGCCCTCTCAGG 1603  
Qy 1580 GCAGCCCTGGCATGAATCCAGACAGACCCACTGCATGCTTTCACCAAGGACATCGCATGA 1639  
Db 1604 GTAGTCCGGG-----CATGCTTCTCCACGTCCACAGGCCA 1639  
Qy 1640 GCCCTGAGTGGCTGCGACGCTCGAATCCACACAGTCACTTTTCCCTGACAGAACT 1699  
Db 1640 GTCTGAGTAGAGATGATCTCCACGCTTACCCCGCCACAGTTCTCTCCAGCGGAAAGCC 1699  
Qy 1700 TGGATTCCTCTGTGGAGTTTGCAGACGA-----CAGGAATTA 1738  
Db 1700 TCCATTACCTGCGACCATGTGTGCACTAGCAGTACGTGTGCAATGGGGGAGCGAGGCCCA 1759  
Qy 1739 GCCATAGTTATACCAACAGCTCCCTCAATGACATTCAGGCTTCAGCGAGGAGGCGACGGG 1798  
Db 1760 ACCATGTGTACACAGCATGTCCTTAATGCTTTTCAAGGCGCTTACGGAATGGCATGTGGCG 1819  
Qy 1799 TCTCATTTAGGGTATCTGTGGCTTTCACACAGACTTAAATATGGCAATTTGCAAAACTCCC 1858  
Db 1820 TAAGCCACGAGCATCTCTGGGATCCCTGATAGAAAGTTAGGGTCCCACTGCATCAT 1879  
Qy 1859 CAGTTAATATGATTCCTCCCTCAGCAAGATGGGAACCTTGGACTCAAAAGACTGTT 1918  
Db 1880 CGGTACAGATTAATCTATCATCTAATTAATTAAGACCTGACTGATCTA-----T 1930  
Qy 1919 TTGACATATATGGGAGCCCTCTGAAGGTACAACTGAGACAGACAGAGAGCTGCATC 1978  
Db 1931 TCGGAGAGCGAAMCAATGACAGACAGGCGACCCCTGGACAGATGAAGGATTTGATCTGC 1990  
Qy 1979 CTGGAAGAGCAAAAGAAACAATGACCCCACTGCCCGCGCGGTGAGCATGTGACAGAG 2038  
Db 1991 CAAGAATGAAAAAG-----GCAACCTCGGGCGATTGCGCAACCTGAGCGGCA 2038  
Qy 2039 CTGACGGGAGAGCAGACATGACAGACAGAAAGGCAACCAACTCTGACAGCTGCTGA 2098  
Db 2039 GTGATCTCTCAAGCCGACCTGCGGACACACAAAGTCACTAACTCTCTCAGCTCTCA 2098  
Qy 2099 CCACCAATCTGATCAGATGGAG-----CCCTCGCCCTTAGCCAGCTCTTTGTGCG 2149  
Db 2099 CCACCTAAATGAGCCTATGATGACCACTTACCCTCCATATGGCGGGTGGAAACCTG 2158  
Qy 2150 ATACAAACAAAGACTCCACAGGTA-----GCTTGCCTGGTTCTGGGCTTACAC 2197  
Db 2159 GGTGTAAAGCGCGGTGCGAGGGAATGTGTCATGAGGTTCTCAAGGGGGAACGGGTCA 2218  
Qy 2198 ATGAAGACTGCTCAAGGAGAAAGATTAATTTTGCACAGACTTGTGACAGAGCACT 2257  
Db 2219 ATGCTACTCTCTGAAAGGAAACACAAATCTCTTCACAGACTCTCTCAGAAACGCTCGT 2278  
Qy 2258 CCCCTGTGACTTGGCCAACTTAACAGCAGAACGACAGGCAAGGCAAGCTTGGCAGAGT 2317  
Db 2279 CGGCTGTGATTTTGGCCAACTTACAGCTGAGGCAAGGGAAGAGCTGTGCAGATTT 2338  
Qy 2318 ----CGAGCAGACAGCTCTGTGATCAGAACTGATTAATAAGAGCCGCTGAGCCCA 2374  
Db 2339 CTGCTGGCGGAGCGCGGTGCGAGTTGGCATTAAACAAGGCTGTCACTGCCA 2398  
Qy 2375 AGAAGAAAGAGAGCACTACTTGGTATTTGGTAGATAAGATGATTAAGATTTG 2434  
Db 2399 AGAAAAAGGACATGCTTGTGCTGCTACTACTGATTAAGATGACAAI----- 2448

---

Qy 2435 GTTTACCAAAATTAACCCCAACTTGAAGAGACTGAGACAGTAAAGACAGATCTGCCATTA 2494  
Db 2449 -----GTACTCAGAGGTTAAAGCTTTTAATAATGAGCGCTGGAGAGA 2488  
Qy 2495 ACACAAATTAATATGCAATGAAGAACTGAGAGAGAGATGAGCTTTGAGCTGTGGACC 2554  
Db 2489 TCAGAGAGAGAGGGGGCAAGATGACTGGGATTAAGCCAGAAATACTGATGGGTCTAGG 2548  
Qy 2555 AGCCTGGCAGTGAAGCTGCAACTTGGAGAGATTTTGGATGTTTGCAGAAATAGTCAAT 2614  
Db 2549 ACCGTGTGAGCCGCTCAAGTGAAGTGAATGATATCTCTGATGATTTCTTCAGAAACAGTCA 2606  
Qy 2615 TACCAAGCTTTTCCAGACAGAGGCGCCCTCTGATGACATGTTGACAAAGCAAG 2674  
Db 2607 -GCCGGGCGCTTCACTGACTCAGGCGCTGTGTCTCTGCTTCTGCTGTGAGCAACAGCT 2665  
Qy 2675 CCATCATCAATGACTTCATATGCAACTCAGAGCTGAAAAACAGCCCTGTCACACCTGTTGGAG 2734  
Db 2666 CAATCATCAATGACATCTGACAGATGACAGAGAGAGCGGAGCTAACATGATGCTCAAC 2725  
Qy 2735 CCCGAAAAACAGCAGCTG---CGAATTTACAGACACTTTTAATAACCCACGACAGGGC 2791  
Db 2726 AGCAGAGAGCACTGCAAGTCAAGTCAAGCCACAGAAATTTCCCTGCTCCAGCGCCGGTTC 2785  
Qy 2792 AACTGGGCAAGTTATTTGCCAAACCAAGAAATTTACACTTGACATACATTCACAAAGCCCA 2851  
Db 2786 AGCCTGCGAGAGCTCTCCGGGTGAGAAAGCGGTGCTCTGACATGAAACATGCTCTTAAG 2845  
Qy 2852 CTGTGCTGAGCTTTTCCCAACATCAGAAACAGTACTGCTTACTGATGATACCTGACG 2911  
Db 2846 GCCCTCAGGCTCAGTATCTTATGATGAGAAACCAACCCCTCACTCAATATGACACAC 2905  
Qy 2912 CAGGAATGATGGTAACTGCTTCTCGGCTTACTATGCCATCTGAGAGATGGAGGCAACGC 3031  
Db 2906 AAGGCAATGATGGAAACCAAGCGCGTATGCGCCATATAGCCCAATGATGTGAAG----- 2959  
Qy 2960 CAGAGATCATGGGTAAATTAATGATGCCGTGATGAGCATGACAGAGCGTGGGGTGTCTC 3019  
Qy 3032 AGAGTTGCGCTGTGAGAGTCACTGTGCTTACCAACAGATGCAATGAACCGG----- 3085  
Db 3020 AGGCTACCGTGGCAGCGCAGCTTCACTGCACTGCTCCGCTGGCGCAGCACTCTTC 3079  
Qy 3086 -----CAGTCCAAGAGATATGATTTGGAACCCAGCAGCAGCATCCCATGA 3133  
Db 3080 TCCCGAGAGGGGCACTGCAACAGCGCATGTGTGCCAAACCTGCGACAGGAATGCCATGA 3139  
Qy 3134 GGCCAGCAGCAGCCTGGCCAAAGACAGCCTTCAAGTCAAGTCAATGATATATAGGGC 3193  
Db 3140 GACCCAAACAGCCGCTGAGCCAGACCAATGCTGCAACCGCAGATGAGGCCAATGCTC 3199  
Qy 3194 CATCTAATTAAGATGAACATGGGGGACCTCAGTATAGCCAAACAAGCTCTCCCA 3253  
Db 3200 AGTCAAAATGAGCATGGGATGGATGGCTGCGACAGTATCCCGACAAACAGCCGCCA 3259  
Qy 3254 ATCAGACTGCCCATGGCTGAAAGCATCTGCTTATAGCAAGGCGCTTTTGGCAGCC 3313  
Db 3260 ATCAGAGGCTCATGATGGCCAGACAGTGTATGCTTATAGAGCGCGGTGGGAAC 3319  
Qy 3314 AAAACAGGCACTTTGGCAGTTCTCAGATCACTTGTCTATGTCACATCTGCAAGCTG 3373  
Db 3320 AGAGCAGTTCCATGTACGGGGGCTTCAGAGAGATGTGTGTGCC-----TCTGCGCTG 3376  
Qy 3374 AGTCTCGAGTGAAGGAGCTCTCTGAGCAAGCTGATATCTGCGGCTTTGGGAATTTG 3433  
Db 3377 AAGGTCGGCTGATGAGGGGGCTTTTACTCAGCCAGCTTACTGCGCCCTCAAAATTTTG 3436  
Qy 3434 ATGGCTGAGAGATTTGATAGAGCTTATAGAAATACCGCACTGTGACAGCAGAGCAAG 3493  
Db 3437 ATGCTGTGAGAGATTTGACAGCTGTGCGCTTGGCATTTCCAGCGCTATGTGGAGCAGGCTCAGC 3496

Y	3494	CAGTGTATCCAGAAAGATTCCTCAATCTCGAGATATCCAAATCATGCTGGAGCAGAAAGGGC	3553
D	3497	CTGTGGAGCAGAGATCAGTTCCCTGGTCAAGACCTTAGTATGTATGTATGATCAGAAAGCCAC	3556
Y	3554	CCGTTTTC-----CCACAGCAGTATGCATCTCAGAGCAAAATGGCCAGGGTAGCTATT	3607
D	3557	CGATGTACGGGGAGCAGAACAGTATGCTGCTGCTCCCGGCAAAACATGACGAGAGCTGGATACA	3616
Y	3608	CTCCCATCGAAGATCCAAACTTTTCACACC-----ATGGGACAGCGGCTTAGTT	3655
D	3617	CGCCAAATCAGAGATCCACTTTTCCACGGCTTCGAAAGCAGAGATGGGACAGAGACCAGAT	3676
Y	3656	ATGCCACACTCGTATGGAGCCAGACCGGGGCTGAGGCCACAGGGGCTTAGTGCAGAAC	3715
D	3677	ACCCATGCTCCGAATGGAGGAGGAGACCGGGAGCTAGACCCGGTGAAGTGTGCTCCAAATC	3736
Y	3716	AGCCAAATCAACTAAGACTTCAACTTTCAGCATCGGCTCCAAAGCAGCAGATCCGACG	3775
D	3737	AACCCAAATACATCAGATGAGTGCAGCTTTCAGCAGACGGCTTCAAGCAGCAGAGATCTCAGC	3796
Y	3776	CACTTATGATCAATCAATCAGCAATGTTTCCATGTGAACTTGACTCTGAGGCTGGAGTAC	3835
D	3797	CTATGATGATCAATCAGATGGGTGGAGTGCATCAATATGAACTTCACTCTCAGACCACGTTTC	3856
Y	3836	CAACACAGGCACTTTATGTCAGCAGATGCTGCGCCACAGACAGAGGGAATCTGAAAC	3895
D	3857	CAAAATCAGGGGACAAATCAACGCTCAGATGCTGCGCCACAGGCGAGAGAGATGTTGCTGACTA	3916
Y	3896	AGCATCTTCGACAGAGCAAAAT-----GCATCAGCAACAGCAAGTTTCAGC	3940
D	3917	ATCATTTTACGTACAGGGCAGCTCGACCGACAGCGCAGCAACAGCAACAGCAGCAGCAGC	3976
Y	3941	AACGAACTTTGATGATGAGAGACAGAGGCTTGAATATGACACAAAGCATGTGTG--GCTC	3997
D	3977	AGCGAGTCTGGCCATGCGGGGCGAGGGGCTCACCATGCAACCAACATGGGGGAGCGG	4036
Y	3998	CTATGTTGATGCCAGCAACTATGAGCAACCTCGGATTTCCCGAGCAATGACACAGCAT	4057
D	4037	CAGCGGGCATGCGAGGACCACTGAGCAACCGCGGATCCCGAGGCAACCTTCAGCAGT	4096
Y	4058	TTCCATTTTCGTCGAAGTACGGAATTAAGTCAAGCAACCTGATTCAGCGCTTTCAGGGGCTA	4117
D	4097	TCCCTTACCTCTCCCACTACGATACAGAGACTGGCTCCCGGCTTCCTCCACACGCCCTT	4156
Y	4118	CGACTCCCCAGAGCCCACTTATGTCACCCCGAATGGCAGATACAGAGATCCCATG	4173
D	4157	TCTCCCGCGGTTCACACCAACTCCTCGGCTTCAGAGCTCTCTCCACAGCGGCATG	4212
RESULT 13			
AC091288_0			
WPCOMMENT			
Sequence split into 4 fragments LOCUS AC091288 Accession AC091288			
Fragment Name Begin End			
AC091288_0 1 110000			
AC091288_1 100001 210000			
AC091288_2 200001 310000			
AC091288_3 300001 404593			
LOCUS AC091288 404593 bp DNA Linear HTG 25-JAN-2002			
DEFINITION MUS MUSCULUS chromosome 2 clone Rp23-254G2 strain C57Bl6/J, ***			
FEATURES			
VERSION AC091288.4 GI:17488573			
KEYWORDS HTG; HTGS; PHASE1.			
SOURCE house mouse.			
ORGANISM Mus musculus			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
1 (bases 1 to 404593)			
Montgomery,K.T., Grills,G., Li,L., Chiu,D., Decker,J., Fustina,M.,			
Goltz,J., Halder,A., Hall,L., Han,J., Ioshikhes,I.P., Lee,E.,			
Peeters,A., Shim,C., Thomas,E. and Kuchnerlapati,R.			
High Throughput Mouse Sequencing			

JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Unpublished  
2 (bases 1 to 404593)  
Montgomery,K.T., Grills,G., Li,L., Chiu,D., Decker,J., Fusina,M.,  
Goltz,J., Haider,A., Hall,L., Han,J., Ioshikhes,I.P., Lee,E.,  
Pereira,A., Shih,C., Thomas,E. and Kucherlapati,R.  
Direct Submission  
Submitted (13-APR-2001) Department of Molecular Genetics, Albert  
Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,  
Bronx, NY 10461, USA  
On Dec 11, 2001 this sequence version replaced gi:13786096.  
-----  
Center: Harvard Partners Genome Center  
-----  
Genome Center  
Center Code: HPGC  
Web site:  
<http://wchanning.bwh.harvard.edu:9088/hpcg/jsp/hpcg/Sequence/mouse.html>  
Contact: gntkm@capecod.bwh.harvard.edu  
-----  
-----Summary Statistics  
Center project name: AEO  
Sequencing vector: pUC18; L08752  
Chemistry: Dye-terminator Big Dye; 100%  
\*Consensus quality: 379539 at least Q20  
\*Consensus quality: 369063 at least Q30  
\*Consensus quality: 353459 at least Q40  
Estimated insert size: agarose-FP - N/A  
\*Estimated insert size: 403233 - sum-of-contigs  
Quality coverage: agarose-FP - N/A  
Quality coverage: 4 x in Q20 bases; sum-of-contigs estimation  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 69 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\*  
1 24548: contig of 24548 bp in length  
\* 24549 24568: gap of unknown length  
\* 24569 40884: contig of 16316 bp in length  
\* 40885 40904: gap of unknown length  
\* 40905 60437: contig of 19533 bp in length  
\* 60438 60457: gap of unknown length  
\* 60458 78011: contig of 17554 bp in length  
\* 78012 78031: gap of unknown length  
\* 78032 94091: contig of 16060 bp in length  
\* 94092 94111: gap of unknown length  
\* 94112 108723: contig of 14612 bp in length  
\* 108724 108743: gap of unknown length  
\* 108744 127422: contig of 18679 bp in length  
\* 127423 127442: gap of unknown length  
\* 127443 142815: contig of 15373 bp in length  
\* 142816 142835: gap of unknown length  
\* 142836 156393: contig of 13556 bp in length  
\* 156394 156413: gap of unknown length  
\* 156414 169259: contig of 12846 bp in length  
\* 169260 169279: gap of unknown length  
\* 169280 181875: contig of 12596 bp in length  
\* 181876 181895: gap of unknown length  
\* 181896 191712: contig of 9817 bp in length  
\* 191713 191732: gap of unknown length  
\* 191733 203964: contig of 12232 bp in length  
\* 203965 203984: gap of unknown length  
\* 203985 212621: contig of 8637 bp in length  
\* 212622 212641: gap of unknown length  
\* 212642 226129: contig of 13486 bp in length  
\* 226130 226149: gap of unknown length  
\* 226150 231591: contig of 5442 bp in length  
\* 231592 231611: gap of unknown length  
\* 231612 241261: contig of 9650 bp in length  
\* 241262 241281: gap of unknown length  
\* 241282 247444: contig of 6163 bp in length  
\* 247445 247464: gap of unknown length

```

* 247465 256788: contig of 9324 bp in length
* 256789 256808: gap of unknown length
* 256809 266558: contig of 9750 bp in length
* 266559 266578: gap of unknown length
* 275232 275232: contig of 8654 bp in length
* 275233 275252: gap of unknown length
* 275253 281450: contig of 6198 bp in length
* 281451 281470: gap of unknown length
* 281471 289954: contig of 8484 bp in length
* 289955 289974: gap of unknown length
* 289975 294111: contig of 4137 bp in length
* 294112 294131: gap of unknown length
* 294132 297704: contig of 3573 bp in length
* 297705 297724: gap of unknown length
* 297725 301771: contig of 4047 bp in length
* 301772 306666: contig of 4875 bp in length
* 306667 306686: gap of unknown length
* 306887 312494: contig of 5808 bp in length
* 312495 312514: gap of unknown length
* 312515 316460: contig of 3946 bp in length
* 316461 316480: gap of unknown length
* 316481 321689: contig of 5209 bp in length
* 321690 321709: gap of unknown length
* 321710 322020: contig of 311 bp in length
* 322021 322040: gap of unknown length
* 322041 325449: contig of 3409 bp in length
* 325450 325469: gap of unknown length
* 325470 329359: contig of 3890 bp in length
* 329360 329379: gap of unknown length
* 329380 331940: contig of 2561 bp in length
* 331941 331960: gap of unknown length
* 331961 337339: contig of 5379 bp in length
* 337340 337359: gap of unknown length
* 337360 340919: contig of 3560 bp in length
* 340920 340939: gap of unknown length
* 340940 343973: contig of 3034 bp in length
* 343974 343993: gap of unknown length
* 343994 347937: contig of 3944 bp in length
* 347938 347957: gap of unknown length
* 347958 350294: contig of 2337 bp in length
* 350295 350314: gap of unknown length
* 350315 354429: contig of 4115 bp in length
* 354430 354449: gap of unknown length
* 354450 357126: contig of 2677 bp in length
* 357127 357146: gap of unknown length
* 357147 359807: contig of 2661 bp in length
* 359808 359827: gap of unknown length
* 359828 361451: contig of 1624 bp in length
* 361452 361471: gap of unknown length
* 361472 364041: contig of 2570 bp in length
* 364042 364061: gap of unknown length
* 364062 367340: contig of 3279 bp in length
* 367341 367360: gap of unknown length
* 367361 370079: contig of 2719 bp in length
* 370080 370099: gap of unknown length
* 370100 372710: contig of 2611 bp in length
* 372711 372730: gap of unknown length
* 372731 375148: contig of 2418 bp in length
* 375149 375168: gap of unknown length
* 375169 376342: contig of 1374 bp in length
* 376343 376562: gap of unknown length
* 376563 376581: contig of 1096 bp in length
* 376582 377678: gap of unknown length
* 377679 379356: contig of 1678 bp in length
* 379357 379376: gap of unknown length
* 379377 380421: contig of 1045 bp in length
* 380422 380441: gap of unknown length
* 380442 381556: contig of 1115 bp in length
* 381557 381576: gap of unknown length
* 381577 382472: contig of 896 bp in length
* 382473 382492: gap of unknown length
* 382493 382428: contig of 1736 bp in length

```

## FEATURES

```

* 384229 384248: gap of unknown length
* 384249 385484: contig of 1236 bp in length
* 385485 385504: gap of unknown length
* 385505 386750: contig of 1246 bp in length
* 386751 386770: gap of unknown length
* 386771 388546: contig of 1776 bp in length
* 388547 388566: gap of unknown length
* 388567 389709: contig of 1143 bp in length
* 389710 389729: gap of unknown length
* 389730 391387: contig of 1658 bp in length
* 391388 391407: gap of unknown length
* 391408 393781: contig of 2374 bp in length
* 393782 393801: gap of unknown length
* 393802 395091: contig of 1290 bp in length
* 395092 395111: gap of unknown length
* 395112 396821: contig of 1710 bp in length
* 396822 396841: gap of unknown length
* 396842 398110: contig of 1265 bp in length
* 398111 398130: gap of unknown length
* 398131 400076: contig of 1946 bp in length
* 400077 400096: gap of unknown length
* 400097 401359: contig of 1263 bp in length
* 401360 401379: gap of unknown length
* 401380 401674: contig of 295 bp in length
* 401675 401694: gap of unknown length
* 401695 403707: contig of 2013 bp in length
* 403708 404593: gap of unknown length
* 403728 404593: contig of 866 bp in length.

```

```

1..404593
/organism="Mus musculus"
/strain="C57BL6/J"

```

Query Match 17.4%; Score 1071.2; DB 2; Length 110000;  
 Best Local Similarity 91.0%; Pred. No.



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 7, 2002, 10:38:09 ; Search time 50.13 seconds

(without alignments)  
5052.156 Million cell updates/sec

Title: US-09-842-256-2  
Perfect score: 7631  
Sequence: 1 MSGMGENTSDPSRAETRRKK.....NOLPGMDIKQEGDTTRKKT 1464

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 segs, 172994929 residues  
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Printing list 45 summaries

Database :

1: SPREMBL\_19: \*  
2: sp\_archaea: \*  
3: sp\_fungi: \*  
4: sp\_human: \*  
5: sp\_invertebrate: \*  
6: sp\_mammal: \*  
7: sp\_mhc: \*  
8: sp\_organelle: \*  
9: sp\_phage: \*  
10: sp\_plant: \*  
11: sp\_rodent: \*  
12: sp\_virus: \*  
13: sp\_vertebrate: \*  
14: sp\_unclassified: \*  
15: sp\_virus: \*  
16: sp\_bacteriap: \*  
17: sp\_archaeap: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match length	DB ID	Description
1	4248	55.7	1505 13 098TW1	Q98TW1 Brachydanio
2	2736	35.8	1420 4 09UPG4	Q9UPG4 homo sapien
3	2734.5	35.8	1417 4 09UPG7	Q9UPG7 homo sapien
4	2729	35.8	1424 4 09Y609	Q9Y609 homo sapien
5	2711	35.5	1412 4 09UPC9	Q9UPC9 homo sapien
6	2688.5	35.4	1391 13 057539	057539 xenopus lae
7	2511	32.9	1398 11 009000	009000 mus musculu
8	2175	28.5	1441 4 043792	043792 homo sapien
9	2174	28.5	1399 4 043793	043793 homo sapien
10	2159	28.3	1441 4 000150	000150 homo sapien
11	2137	28.0	1405 11 070366	070366 mus musculu
12	2108	27.6	1405 11 061202	061202 mus musculu
13	2106	27.6	1447 11 070365	070365 mus musculu
14	1684	22.1	1082 11 09EP02	09EP02 rattus norv
15	1033.5	13.5	1061 4 013420	013420 homo sapien
16	545	7.1	416 4 013071	013071 homo sapien

17	508	6.7	203	4	Q9BR49	Q9BR49 homo sapien
18	495.5	6.5	2035	5	Q9GS19	Q9GS19 drosophila
19	425	5.6	1778	8	Q9VLDB	Q9VLDB drosophila
20	374	4.9	326	4	015406	015406 homo sapien
21	340	4.5	2232	5	P91365	P91365 caenorhabdi
22	340	4.5	2302	5	Q9N693	Q9N693 drosophila
23	339.5	4.4	893	13	Q9W6J4	Q9W6J4 Brachydanio
24	339	4.4	2310	5	Q9RA9	Q9RA9 drosophila
25	334	4.4	5476	5	Q9N117	Q9N117 drosophila
26	334	4.4	5533	5	Q9VPL2	Q9VPL2 drosophila
27	334	4.4	5534	5	Q9NHN1	Q9NHN1 drosophila
28	334	4.4	5280	5	Q9VPL1	Q9VPL1 drosophila
29	332	4.4	2280	5	Q9V8E6	Q9V8E6 drosophila
30	328	4.3	862	11	Q9WVS9	Q9WVS9 rattus norv
31	328	4.3	865	11	Q91XB0	Q91XB0 spalax gall
32	328	4.3	5533	5	Q9V6C3	Q9V6C3 drosophila
33	325	4.3	865	11	Q91YB2	Q91YB2 spalax carm
34	325	4.3	865	11	Q91YA8	Q91YA8 spalax juda
35	321.5	4.2	778	13	Q9PUA3	Q9PUA3 xenopus lae
36	321	4.2	825	13	Q91906	Q91906 xenopus lae
37	319	4.2	875	13	Q9W7C3	Q9W7C3 gallus gall
38	318	4.2	853	13	Q9W602	Q9W602 gallus gall
39	314.5	4.1	870	6	Q9X7A4	Q9X7A4 bos taurus
40	314.5	4.1	2703	5	Q9VEG7	Q9VEG7 drosophila
41	313.5	4.1	4957	4	014687	014687 homo sapien
42	313.5	4.1	5262	4	014686	014686 homo sapien
43	312.5	4.1	832	11	Q920T1	Q920T1 rattus norv
44	310.5	4.1	852	13	Q91A11	Q91A11 gallus gall
45	309	4.0	2715	5	061603	061603 drosophila

## ALIGNMENTS

RESULT 1  
ID 098TW1 PRELIMINARY: PRT; 1505 AA.  
AC 098TW1:  
DT 01-JUN-2001 (TREMBLrel. 17, Created)  
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE TRANSCRIPTIONAL INTERMEDIARY FACTOR 2.  
OS Brachydanio rerio (zebrafish) (zebrafish).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;  
OC Cypriniformes; Cyprinidae; Danio.  
OX NCBI\_TaxID=7955;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Tan J.H., Kara A.H., Chan W.K.;  
RT "Molecular cloning and characterization of zebrafish nuclear receptor  
coactivator TIF2.";  
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF123989; AAK11606.1; -  
DR InterPro: IPR001092; HLH\_dim.  
DR InterPro: IPR000664; PAS.  
DR Pfam: PF00989; PAS; 1.  
DR SMART: SM00353; HLH; 1.  
DR SMART: SM00091; PAS; 1.  
SQ SEQUENCE 1505 AA; 161793 MW; 88BD0A8435B14497 CRC64;

Query Match  
Best Local Similarity 55.7%; Score 4248; DB 13; Length 1505;  
Matches 900; Conservat 100; Mismatches 310; Indels 154; Gaps 36;

Qy 1 MSGMGENTSDPSRAETRRKCEP-DOLGSPKRNTERKRNREDENKYIEELAIPIAND 59  
Db 1 MSAVENSDDPSRAEQRRKESPSDLGSPKRNTERKRNREDENKYIEELAIPIAND 60  
Qy 60 IDNFNFKDKCAILKETVKQIROIKEOKAANAANDEVOKSDVSTGGGVIDKDALGPM 119  
Db 61 IDNFNFKDKCAILKETVKQIROIKEOKAANAANDEVOKSDVSTGGGVIDKDALGPM 120





```

Db 298 RCIQREFSINDQSWMS-QKRYOEAFLNGHAEPTVYRESLADGTTVTAQTKKLEFPNV 356
QY 361 TNEPOLVLSHMLHREQNVNVPMDLITGTMGKPLNFSSNSPAHQALCSGNFGODMTLS 420
Db 357 TNRHGVSTHFLQRENGVRPNPNVYGQIRPFMAGCNSVGG----NMSPNQGLOMP 412
QY 421 SNINFPINGKEQMGMPMGFRGSGGNNHVS---GMOA-TTPGGSNVALKMSNPSSSPG 476
Db 413 SSRAYGLADSTTGOMSGARYGSSNIALTPPGMQSPSSYONNNNGLMSSPPhSPG 472
QY 477 MNGOPTMSLSPHBMSPGVASPRTPPSQFSPAGSLHSVVGCCSSNGSHSTNSLNA 536
Db 473 LADNOONIMTSPKRR-----GSKTASHQFSPVAGVHSMASGNGN-HSFSSSLSLA 525
QY 537 LQALSEGHVGLSSSLASPDLMKGNLQNSPVNNMPPLSKMGSLSKDFGLXGE--PSE 594
Db 526 LQALISEGVGLSTLSTLSSPGPK---LDNSP-NNNITQPSKVSNOQDSKSPGLFCODQDPE 581
QY 595 GTTGOAESCHGPEQKETTNDPNLEPPAVSSERADQSKRLHDSKQTKLLQLLTKSD--- 650
Db 582 SSMQCSNDRDLSD-KESKES-----SVGAENQGRPLESKGHKKLLQLLTCSDDRGH 634
QY 651 -QMEPSPSLASL-----SDTNKDSVGLSPGSGSTHGTSLKEKHKLTHRLDQSSS 699
Db 635 SSLTNSPLDSSCKESSVYSPSGVSSSTSGVSTSNMHSGLQEKHRLHLKLGONS 694
QY 700 PVDLAKLTATGCKDLSESSSTAPGSEVYIKOEPVSPKKE-NALLRYLLDKDQTKD 758
Db 695 PAEVAKTTAATGKD---TSSITSCGCGNVYKQEQLSPKKENNALLRYLLDDBDPSDAL 751
QY 759 LPEITPKLERLSDKTDPASTKLIAMKTEKE-EMSEFGDQPSGLDNEILLDQNSQ 817
Db 752 SKELQPOVEVDNMKSQCTSTTIPSSQEKDPKIKTETSESGSDDLNDLAIIDLTTSSD 811
QY 818 LPQLFPTRRGAGASVDKAIINDLMQLAENSPTVPAQKTAALISOSTFNPNRPQ 877
Db 812 FYNNSSISN-GSHLGT-KQOVPQ-----GTNSLGLKSSQSV-QSIRP- 850
QY 878 LGRLLPNQNLPLD--ITLQSPGAGPPPIRNSSPYVIP-QPGMNGNOQMIGN 934
Db 851 -----PYNRVSLDPSVSGSSPYKNTSAFPMRLKQPMLGSGNPMMSQENYGS 900
QY 935 SSTGMIGNSASRPTMSEGMA-POSSAVRYTCAATTS-----AMNRPVOGGMIRN 983
Db 901 SMGGPNNNTVYTOTPSSGDMGLPNSKAGRMEMNSNMGPRGDDYNTSLRPAAGSI-- 958
QY 984 PASIIPRPSSQGOQRTL--OSOVNIGSPSELMNMGSGQYOSQQAAPNQTAPWPSIL 1041
Db 959 --PILPILRNSIFPARVYLOOQOQMLQMPGELPMGGMANPYG-QAALNQDLSGWPDMGL 1015
QY 1042 PIDQASFASONRQPFSSPDLLCPHRAESPDEGALLDQLYLALRNF--GLIEIDRA 1099
Db 1016 SMEVSHGCTGNRLPLRNSLDDIVGPRNSLGEQSERALLDQHLHTLSNTATGIEIDRA 1075
QY 1100 LGIPELVSGQAVDPPE--FSSQDSNIMLQKAPVPOQYASQAKMAQSGYSTPMQDENH 1157
Db 1076 LGIPELVNCOALEPKODAFQGOEAAVMMQKAGLYGQTPAOGPPMQG-----GFH 1127
QY 1158 TMOGRPTATLRMQ-----PRGLRPTGLVQ---NONNOLRLQLOHLQAO---ON 1202
Db 1128 LOGOSPEFNSMMQNOQGFPLQGMHPRANIRPTNTKOLRMQLOQRLQOQFLNQS 1187
QY 1203 RQPLANOISN-VSNVNLTLRPV-PTQAPINAOMLAQOREILLNOHLR----- 1248
Db 1188 RQALTELKMEKPTAGGAVMRPMQPOGFLNAQVAVARSSELLSHNRQGRVAMMMQOQ 1247
QY 1249 -----QROMHQOQVOQRTLLMRGQGLMTPSNVAAPSMPATMSPRILPOANAQOFP 1304
Db 1248 QOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 1307
QY 1305 YGISOQDPDPTGATTPQSLPMSPMAHNTQSPMMQOQOQAPQAPQADINDINGMAQNGN 1364
Db 1308 YGMQOQDPDPAFGKYSSPPNMMSSRMGPSONPMQHPQOASITQY--SEKGMWPSGNLARN 1366

```

```

QY 1365 SMFSQSPPHFGQOANTSMYSNNMNINVSMTATNGMSSNMOTGOISMTSVTSVTSGL 1424
Db 1367 SSFSQO---QFAHQGNPAVYS-----MVHNGSSGHMGQNNMNMPSMGM 1408
QY 1425 SSMGPEQ 1431
Db 1409 -PMGPDQ 1414

RESULT 4
09Y609
ID 09Y609 PRELIMINARY; PRT; 1424 AA.
AC 09Y609;
DT 01-NOV-1999 (TREMblrel, 12, Created)
DT 01-NOV-1999 (TREMblrel, 12, last sequence update)
DT 01-JUN-2001 (TREMblrel, 17, last annotation update)
DE THYROID HORMONE RECEPTOR ACTIVATOR MOLECULE (DJI04G16.2) (NUCLEAR
DE RECEPTOR COACTIVATOR 3 (THYROID HORMONE RECEPTOR ACTIVATOR MOLECULE
DE TRAM-1, RECEPTOR-ASSOCIATED COACTIVATOR RAC3, AMPLIFIED IN BREAST
DE CANCER AIB1, ACTR)).
GN TRAM-1 OR NCOA3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98010595; PubMed=9346901;
RA Takeshita A., Cardona G.R., Kolbuchi N., Suen C.S., Chin W.W.;
RT "TRAM-1, A novel 160-kDa thyroid hormone receptor activator molecule,
RT exhibits distinct properties from steroid receptor coactivator-1."
RL J. Biol. Chem. 272:27629-27634(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA Skuce C.;
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF016031; AAC51849.1; -.
DR EMBL; AL034418; CAC17693.1; -.
DR InterPro: IPR001092; HLH_dlm.
DR InterPro: IPR000014; PAS.
DR Pfam: PF00989; PAS_1.
DR SMART; SM00353; HLH; 1.
DR SMART; SM00091; PAS; 1.
KW Receptor.
SQ
SEQUENCE 1424 AA; 155293 MW; 732CDF0423161679 CRC64;

Query Match 35.8%; Score 2729; DB 4; Length 1424;
Best Local Similarity 42.5%; Pred. No. 3e-158;
Matches 644; Conservative 243; Mismatches 451; Indels 176; Gaps 44;

QY 1 MSGMGENTSPSAPAEFRKREKCPDQIGSPKRTKERNRROEKYIEELAELEAFANPNDI 60
Db 1 MSGGEML-DPLASDSKKRLPCDTPGQGLTCSGKRRRQESKYIEELAELEASLSDI 59
QY 61 DNEFNKDKCAILKETVQKQIKTEQEKAAANIDEVQKSDSVSTGGVYIDKDALGPMML 120
Db 60 DNEFNKDKCAILKETVQKQIKTEQEK--TISDDVQKADVSTGGVYIDKXSDPLLL 118
QY 121 EALDGFYVNLGNNVYVSSENTQYIKRYQOELANKSVYSLHVGDTHEFYVNLPKSI 180
Db 119 QALDGFYVNLGNNVYVSSENTQYIKRYQOELANKSVYSLHVGDTHEFYVNLPKSI 177
QY 181 VNGSGSGEPPRRNSHTFENRMLVYKPLPSEEEGHNOQEAHOKYEMOGCPAVSOPSIKE 240
Db 178 VNGSWTNEQKQKSHFNCRLMLKTPHDILDINSPKRRQRTETMOCFALSOPRAME 237
QY 241 EGEDLQSLICIVARRVPMKEPVLPSSESFTTQDLOGKITSDTSTRAAMPGEVDLY 300
Db 238 EGEDLQSCMIVARRITTGRTTPSNPESFTTHDLDSGVYVNDTSLRSMSMPGEEDII 297
QY 301 RCIQFHAQHEGESVYAKRHHHEVYLRQGLAFSQTLYRSLSDGTLVAAQTKKILRSQT 360

```

[illegible]

Db	1308	QPPYPTNMGQDQDPRAAGRVSSPPNNAMSSMGSSQNPAMQHPDAASTIYOS-SEMKGP	1366
Qy	1358	QGNMGNSMFSSQSDPBFHGOQANTSMYSNNMNI NVSMATNTGSMSSMNOQTGQISMTSVT	1417
Db	1367	SGNLARNSSEFSQO--QFAHQGNPAVYS-----AVHNMGSGSHMGOMNMN	1408
Qy	1418	SVSTGSLSSMPREQ 1431	
Db	1409	PMPMSGM-PMGPQD 1421	
RESULT	5		
Q9UPC9			
AC	Q9UPC9	PRELIMINARY;	PRT; 1412 AA.
AC	Q9UPC9;		
DT	01-MAY-2000 (TREMBLrel. 13, Created)		
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)		
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)		
DE	NUCLEAR RECEPTOR COACTIVATOR.		
GN	ACR1.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.		
OX	NCBI_TaxID=9606;		
KN	[1]		
RX	SEQUENCE FROM N.A.		
RX	MEDLINE=97410321; PubMed=9267036;		
RA	Chen H., Lin R.J., Schaltz R.L., Chakravarti D., Nash A., Nagy L.,		
RA	Privalsky M.L., Nakatani Y., Evans R.M.		
RT	"Nuclear receptor coactivator ACR1 is a novel histone		
RT	acetyltransferase and forms a multimeric activation complex with P/CAF		
RL	and CBP/p300."		
DR	Cell 90:569-580(1997).		
DR	EMBL: AF036892; AAB92368.1; -		
DR	TRANSPAC: T04640;		
DR	InterPro: IPR001092; HLH_dlm.		
DR	InterPro: IPR000014; PAS.		
DR	Pfam: PF00989; PAS; 1.		
DR	SMART: SMO0353; HLH; 1.		
DR	SMART: SMO0091; PAS; 1.		
KW	Receptor.		
SO	SEQUENCE 1412 AA; 154115 MW; 84573E2F383399D0 CRC64;		
Query Match	35.5%; Score 2711; DB 4; Length 1412;		
Best Local Similarity	42.3%; Pred. No. 3.8e-157;		
Matches 642; Conservative 242; Mismatches 439; Indels 194; Gaps 46;			
Qy	1 MSGCGENTSDPSRAETRRKRECEPDOLGSPKRNTERKRNREDENKYIEELAEIIFANPNDI	60	
Db	1 MSGLEENL-DPLASSRRKRLCPDTPGGGLTCSGEKRRREDSKYIEELAEIISANLSDI	59	
Qy	61 DNFENKPKCALIKETVROIROIKROEKRAAANIDEVOKSDVSSNGOCYIWDALGPMML	120	
Db	60 DNFENKPKCALIKETVROIROIKROGK-TISNDVDQKADVSSGQGVIDSDSGPLL	118	
Qy	121 EALDQFFEVNLEGNVFESENVTOYLRYNOBELANKSYSLIHYGDTHFEYKNIPLKSI	180	
Db	119 QALDQFLFVYVNDGNIVFSENVTOYLOKQDDLYNTSYNLIHDEKRDPLKN-LPST	177	
Qy	181 VNGGWSGEGPPRRNSHTENCRLVLRPLDSEEGHDNOEAHQKYETMOCFAVSQPSIKE	240	
Db	178 VNGWSWTNETQOKSHTEFCRLMKTPHDILEDINASPERRQRYETMOCFALSPRAAME	237	
Qy	241 EEDDQSLCLICARVMPKERVYLSSSEFTTROLQKITSLDSTIRAAKPPWEDLY	300	
Db	238 EEDDQSLCLICARVITTGERTFPNPSBFIRHDLGKVVINDIYNSLRSSMRPFEDI	297	
Qy	301 RRCIDKFNAHQHGESEVSAKHHHHVLRQL-----AFSQIYRFSLDGTLVAAQ	350	
Db	298 RRCIDKFVFSLSGQWS-QKRHYQEVTSGLISPAIYLANGAEETVYRFSLDGTLVTAQ	356	
Qy	351 TRSKILRSQTTDEPOLVLSLHLRHREONCVANPDLTGQTMKPLNPISNSPAHQALCS	410	

```

Db 1352 GWSGSLNARNSFSQO--QFAHOGNPAVYS-----MHMGSSGHMGOM 1393
Qy 1415 SVTSYSTGLSSKMGPEO 1431
Db 1394 MNMMPMNSGM-FMGPDQ 1409

RESULT 6
057539 PRELIMINARY: PRT: 1391 AA.
AC 057539:
ID 057539:
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DE 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE RETINOID X RECEPTOR-INTERACTING COACTIVATOR XSRC-3.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae;
OC Xenopodinae; Xenopus.
OC NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RA Kim H.-Y., Lee S.-K., Na S.-Y., Choi H.-S., Lee J.W.;
RT "Molecular cloning of xSRC-3, a novel retinoid x receptor-interacting
RT coactivator from Xenopus, that is related to AIB1, p/CIP and TIF2."
RL Mol. Endocrinol. 0:0-0(1998).
DR EMBL: AF044080; AAC12927.1; -.
DR Interpro: IPR001092; HLH.dlm.
DR Interpro: IPR000014; PAS.
DR Pfam: PF00989; PAS; 1.
DR SMART: SM00353; HLH; 1.
DR SMART: SM00091; PAS; 1.
DR Receptor.
KW SEQUENCE 1391 AA; 152532 MW; AD28F5CD934AC33D CRC64;
SQ

Query Match 35.4%; Score 2698.5; DB 13; Length 1391;
Best Local Similarity 43.3%; Pred. No. 2.1e-156;
Matches 648; Conservative 239; Mismatches 436; Indels 173; Gaps 51;

Qy 1 MSGGENTSDPSRAETRRKK--ECPPDLGSPPKRNRKREDEKRYIELAELIPANFN 58
Db 1 MSGLGENSELDPLASETRKRKPPSC-DTPGGDLTCSSEKRRREDESYIEELADLIANIS 59
Qy 59 DIDNFKPKDKAILKETVQKIQIKOEKAAANIDEVOKSDVSSTGGGVIDKDALGPM 118
Db 60 DIDNFKPKDKAILKETVQKIQIKOEK--ASSNDVDQKADVSSTGGGVIDKDALGPM 118
Qy 119 MEALDGFPPVNLBEGNVFVSENVQYLYRNOELMKNKSVSILHVGHTERVKKLLRK 178
Db 119 LQALDGFVYVNRRESIYFVSENVQYLYRNOELMKNKSVSILHVEDRKFLLK-LRK 177
Qy 179 STYNGSWGSEPPRRSHSFNCRLKPLPDSEEGHDNOEAKOKETMQCFVANSQPKSI 238
Db 178 STYNGVPMFSETPRQKSHFNCRLKPLPDSEEGHDNOEAKOKETMQCFVANSQPKSI 236
Qy 239 KEEGEDLQSCILICVARVPKERPVLPSSSEFTTRDLOGKITSLDTSTWRAAMKGWED 298
Db 237 IEEGEDLQSCIMICVARITTAERAFSANPESFTTRDLDLGKVYVINDANSIRSSMRGFE 296
Qy 299 LYRRCIQKFAHOGESVSTAKRHHHEVLRQGLAEFQIYRFSLSDTGLVLAQTKSKLIS 358
Db 297 TIRRCIQRFLEHSEGOPTWY-KRHYDEAVYHGLSETPLYRFSLADGTMYAOTKSKLIFRN 355
Qy 359 QTNEPOLVLSLMLHREOVNVCVANNPDLTGOTGKPLNPISNSPAAHQALCSGNPGQDMT 418
Db 356 PVTNDHGFVSTHFLDRBOGTRPNPNMAQGITRPOANPULPNT-----MNSMPQAMQ 409
Qy 419 LSSNIFPIPKQEQMGMPGRFGSGGM--NHVSGMOATTPQ-GSNYALKNNSPQSS 474
Db 410 -QGNRYMGMDPNSMAQMGMRKYSFGNMAPVQAQVGOQSPYQNNNSNGLNNSPQSS 468
Qy 475 PGANPGOPTSMLSPRHRMSPGAVGSPRIIPSPQSPAGSLHSPVGVCSSTGNS--HSTYNS 532

```

```
Db 469 PCMANNOPLMAYSPRRASPKMAS-----NOFSVPGMNPMG---SSGNAGGSFSS 519
Oy 533 SLNALQALSEGHVSLGSLASPDLMKGNLONSPPVNMNPPRLSKGSLDSKCFGLYGP 592
Db 530 SLNALHAISEGVSSLSLSSPGQKVEN--NS--NMNPMQKICNDCKSPSLGYC- 574
Oy 593 SEGTTGOAESS-CHPEDEKETMDPMLPPAVSSERADGOSRLDSGQKRLDILLTKSDQ 651
Db 575 ----OGVSSSVCOSSGHEHLEKDYKENV-FEGSESQSAESGHHKLLDILLCFTEE 629
Oy 652 MEPSPLASSDPTNKDSTGSLPGSG-----STGCTSLKEKHILHRLLDSS 698
Db 630 RGSISLSSSSMDC-KDSNVITSPSGVSSSTIGVSSSTSLHSGMLQEKHRLHLKLLQNGN 688
Oy 699 SPVDLAKITAEATGKDLQOESSSTAPGSVTTKQEPVSPKKE-NALLRYLDKDDTDKI 757
Db 669 SPAEYAKITAEATGKDLQOESSSTAPGSVTTKQEPVSPKKE-NALLRYLDKDDMDP 748
Oy 758 GLPETPKLERLDSKTDPAASNTKLIAKTEKE-EMSFEPGDPGSELDLLEILLDONS 816
Db 749 LAKDKPVEHMDIKMGSCSSSNVPTSSODKEVKTITEGEEVPGDLDNLIDLGLAGS 808
Oy 817 QLPOLFPDTRPGAPASVDCQAIINDLMQLAENSPTVPGAKTALRISOSTFNNPRPG 876
Db 809 ---DFYSNMSRABLDLPKQVPFQDSPTL-AMRSPDSMGSR----- 847
Oy 877 QLGRLPQNPLDITLTPGAGFPPIRKNSSPVTPQCGMNGMNGMIGQGLGSS 936
Db 848 -----PPFNAMSLDSRSSST-----PPRVNVSFPMPLPKQMGISPPMDOQDNFG- 893
Oy 937 TGMIGNASRP--TWSEMAPOSSAV-----RVTCAAT-----TSAMNRPVCGMIRN 983
Db 894 -VMMSGNRKSNQHPRGDAMONSAVNRLEPPNVGSGRGCRDYSSATRAMGGM- 950
Oy 984 PAASIMPRSSOPGQROTLDQSO--VMNIGPSELENNMGROYSOQOAPRNPOTAPRESIL 1041
Db 951 --PGLTFRNSNIPGSRPVMOQOQHILPMRPNMAMSGSNPYG-QQAPSNPRGSPDAIM 1007
Oy 1042 PILDASFSQNRQRPFGSSPDLLCPHRAEBSDEGALLDLYLALRNF--GLEIDRA 1099
Db 1008 -MNGRGGAQNRKOLRNLIDLCPSTVEGOTDEIALLDHLTLSTWDATGLEIDRA 1066
Oy 1100 LGIPELVASQAVD--PROFSSODSNIMLEOKAPVFPQOYASO-QOMOGSVSPMODNF 1156
Db 1067 LGIPELVASQALEPQDPDSTORQSGRVMIIDQKPRMIGOHYAGGASAGGNNMQ---- 1122
Oy 1157 HTMGORPSYATLR-----MORRGLRP--TGLVQ--NOPNOLRLQHLRLAQ----- 1200
Db 1123 ---GQHPREFNVMGOMNOQOQHNRLOGMHRANLIRPNINIRKOLRMDLQORLQOQFLN 1179
Oy 1201 QNRQPLMNOISVSNVNI--TLRPVYPT----QAPINAOMLAOROEILNOLRQRMHQ 1255
Db 1180 QNRQALEKRVPMRPGAGVAPVMOQTPVSOQGFILNAOVAKNREILISHQIRQRMAM 1239
Oy 1256 QOVQORTLMMGGQGLMTPRNVAVRSGMPTMSNPRIIPANAOQFFPRNTYGISQOPDGEF 1315
Db 1240 MOQOQOQ-----OPQAFSPRPNTYASAKMNPGLGPRMPQAPQOQFSYPRNYGINTDPTF 1295
Oy 1316 TGAETTPSLMSPMAHTQSPMOQSOANPAYQAPSDINGMAGMNGNSMSPHF 1375
Db 1296 GRVSSPRAMSSRABRQNHQPTQMYPS-----PDKKMPSGGMARPNSTPQOQYSH- 1350
Oy 1376 GQOANTSMYSNNMNIINVMATNTGMSMNQNTGOISMTSVTSGLSSGPRQ 1431
Db 1351 --QTNPATY--MMHMN-----GNGNHMGOM-----NINSLPMGSM-PMGPRQ 1388
```

```
RESULT 7
ID 009000 PRELIMINARY; PRF; 1398 AA.
AC 009000;
DT 01-JUL-1997 (TREMBLrel, 04, Created)
```

```
DT 01-AUG-1998 (TREMBLrel, 07, Last sequence update)
DE 01-DEC-2001 (TREMBLrel, 19, Last annotation update)
CN P300/CBP/CO-INTEGRATOR PROTEIN.
OC Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9736097; PubMed=9192892;
RA Torchia J., Rose D.W., Inostroza J., Kamel Y., Westin S., Glass C.K.,
RA Rosenfeld M.G.;
RT "The transcriptional co-activator p/CIP binds CBP and mediates
RL nuclear-receptor function."
RL Nature 387:677-684 (1997).
RN [2]
RP SEQUENCE FROM N.A.
RA Torchia J., Rosenfeld M.G.;
RL Submitted (MAR-1998) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF000581; AAC05020.1; -.
DR TRANSFAC: T04638; Ncoa3.
DR MGD: MGI:1276535; Ncoa3.
DR InterPro: IPR001092; HLH_dim.
DR InterPro: IPR000014; PAS.
DR Pfam: PF00989; PAS. 1.
DR SMART: SM00353; HLH; 1.
DR SMART: SM00091; PAS; 1.
DR SMART: SM00091; PAS; 1.
SQ SEQUENCE 1398 AA; 151573 MW; EF44E92735816C24 CRC64;

Query Match 32.9%; Score 2511; DB 11; Length 1398;
Best Local Similarity 41.2%; Pred. No. 6,4e-145;
Matches 621; Conservative 232; Mismatches 466; Indels 188; Gaps 47;

Oy 1 MSGMGENTSDPSRATRRKKECPDQGPSKRNTERKRNEDQNKYIEELAEILFANFDI 60
Db 1 MSGLESSLDPALASRRKRLPCDAPGQGLVYSGEKKMRDEQSKITEELAEILSNLSDI 60
Oy 61 DNFNFKPKCALIKETVQKIQIKQEKAAANIDEVOKSDVSTGQGVIDKALGPMML 120
Db 61 DNFNFKPKCALIKETVQKIQIKQEKQKTISSD-DDVQKADVSTGQGVIDKSLGPILL 119
Oy 121 EALDFFEVVNLGCVNVVSENVTOYLKRYNOELMANKSYSLIHGHDTPEYKNLPSI 180
Db 120 QALDFFELVNVNDGNIIVSENVTOYLOKQDLVNTSYSLIHPRRRKDFLNTYQNDL 179
Oy 181 VNGSGMSGPPRRNSHTFNCRLVRLPDSSEEGHNDQBAHQYETMOCFAVSOPKSIK 240
Db 180 MEFGLMKTRODKKARYILVRLM-K-THDILEDVNASPETRKQRYETMOCFALSOPRAMLE 238
Oy 241 EGEDLQSLICVARRVPKKEPVLPSSEFTTRDLOGKITSLDTSTRAANKPMEDLV 300
Db 239 EGEDLQCMICVARRV---TAPFPSSPSEFTTRHDLGKVNVVIDNLSRSSMRPFEDI 295
Oy 301 RRCIOKFAQHEGESVSTAKRHHNEVYLRQGLAFSQTIRFSLSDGLVLAQTKSKILRSOT 360
Db 296 RRCIORFESLNDGQWS--OKRHYQEAUVYHGAETFYRFSLADGTIVSAQTKSKILRRPV 354
Oy 361 TNEPQLVSLHLHLHREOVNVANPDLTGOTMGRPLNPTSSNS---PRAHOLCSGNPQD 416
Db 355 TNDRGFISTHFLQREON-----GYRNPPLIPQDGIAPPAAGCVSSHPN 401
Oy 417 MTLSSNINEPIPKQGMKPRGSGGNHVS-GNQAATP---QGSNTALKNMSPQ 472
Db 402 VQMGSRITYGVDPDNTQMGAGARVAGASSVASLIPGQSLGSPSSYQSSYGLSMSPRH 461
Oy 473 SSPGMNPGOPTSMLSPRIRASGVASPRIPPSQSPAGSLSPGVGVSSSTGNSTYNS 532
Db 462 GSPGLGPNQOQNTMISPRR-----GSPKMASHOPSPAGASPMGPGGNTG-SHSFSS 514
Oy 533 SLNALQALSEGHVSLGSLASPDLMKGNLONSPPVNMNPPRLSKGSLDSKDCFCFLYGE- 591
```





```

Db 681 SFLERHKLHRLLOE-GSPSDITTLVSEPPKKD-SASTSVSVTGOVGNSSIKLELDA 738
Qy 728 KKKK---NALLRYLLDKD-----TKDGLPEIPFKLERLSKTDPASTKLIAMKTEKE 789
Db 739 KKKESKHQRLRYLLDKREKDLRSTPNLSLDVKKYKVEKE-QMPCNTNPTPTKPTPE 797
Qy 790 EMSEFPDQPSSELDNLEIIDLON-SQLPOLFPDTPRGAPGVDAKOAIINDIMOLTA 848
Db 798 EIKLEAQSOTFADLDQFQOLLPTLEKAQOLPGLCETDRMD--GAVTSVTIKSEL----- 849
Qy 849 ENSPTPVGAOKTALRISQSTFNNRPGOLRLPNQMLPDLITLOSTPGACPPPIRNS 908
Db 850 -----LPASIQSAT----- 858
Qy 909 SPYSVIPOGMMGNMGIGNGLNSSTGMIGNSASRPTMPSGMAVQSSAVRYTCAAT 968
Db 859 -----ARP----- 861
Qy 969 TSANRPVYOGMIRNPASIFMRSSQPGOROTLOSQVMNIGPSELEMMNGGPQYSSQQA 1028
Db 862 TSRLNR-----LP-----ELELEALDNQFGPRT 885
Qy 1029 PPNOTAMP-ESLPIDQASASQNRQPGSSPDLLCPHPAESPDGALLDOL--YL 1085
Db 886 --GDQIPWTNNTVTAIINQ--SKSEDOICISQIDELLCPTTVEGRNDEKALLDQVSEFL 940
Qy 1086 ALRNPDLGELIDRALGIPELYSQSQAVD--PEQSSQSDN--IMLEOKAPFPOQYASQA 1141
Db 941 SGKQETELAEIDRALGIDKLIV-QGGGLDVLSERPPOQATPPLIMEERNLYSQYSSPS 999
Qy 1142 QMAQSYSPMODPNFHTGGRPSYATLRMPRP-----GLRPTGLYON--QPNOL 1189
Db 1000 PTAN-----LSPRPGQAVRKQPSLCTMPOVYTPRGASRPGMGMPROTILNRPAPAKOL 1054
Qy 1190 RLQLOHRLQAO-----QNRQPLANOISVSNVNTLTPRC--VFQAPINAOMLAQOR 1240
Db 1055 RLQLOQRLOGQOOLIHNRQAQILNOFATAVGIMRSGMOQOLTPROPPLNAQMAQOR 1114
Qy 1241 EILNQHRLOROMHQOQOQVQOATLMMRQO--GLNMPPSVAVASGMAPATSNRPIRPAANO 1298
Db 1115 ELYSQOHRORLOI-----QQRAMLRQOSFGNNLPPS--SGLEVQNGNPRALPGADPQ 1166
Qy 1299 FPPRPNTGISQOPDPGFTG-----ATTPQSPU-----MSPMA 1331
Db 1167 FPPRPNTGNTGTPRASPSPSQAANPEASLARNSVSGKMTGNIGQGTGCTGINTPQMO 1226
Qy 1332 HT-----QSPMAQOSQANPAYQAPSDINGW 1356
Db 1227 QNVFOYRPGAGVPOGEANFAPSLSPGSSMVPMPPIPPQSSILQOTPPASGYQSP-DMKAW 1285
Qy 1357 AOGMMGMSRMSO--QSPPHGQQAQNTSMYNNMNNINVSMTNTNGMSSMNOQTQISMT 1414
Db 1286 QOQAIQNNVNSQAVQONP--TPAQPCVY-NNMSITVSMAGNTNVQNMNPMMAQOM- 1340
Qy 1415 SVTSVSTGSSMSGPEQVNDPALKGNLFPNQLPGMDMIKQECDTTR 1462
Db 1341 --SSLQPMGNTVCPQEPQINDPALNHTGLYCQSLSTDLKTEADCTQ 1386

```

```

RESULT 9
ID 043793 PRELIMINARY; PRT: 1399 AA.
AC 043793:
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE STEROID RECEPTOR COACTIVATOR 1E.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eulheta; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.

```

```

RX MEDLINE=98090466; PubMed=9427757;
RA Kalkhoven E., Valentine J.E., Heery D.M., Parker M.G.;
RT "Isoforms of steroid receptor coactivator 1 differ in their ability to
RT potentiate transcription by the oestrogen receptor.";
DR EMBL; AJ000882; CAA04372.1; -.
KW Receptor.
FT CHAIN 2 1399 POTENTIAL.
SQ SEQUENCE 1399 AA; 152384 MW; DACE967BJ31AC6B69 CRC64;

Query Match 28.5%; Score 2174; DB 4; Length 1399;
Best Local Similarity 36.2%; Pred. No. 2,6e-124;
Matches 581; Conservative 236; Mismatches 422; Indels 368; Gaps 61;

Qy 1 MSGCEMTSDPSRATRRKKECP-DQLGSPKRNTEKRNDEQNKYIELLAELIFANRD 59
Db 1 MSGLCDSSSDPANPDSHKRGSPCDTLAS--TEKRRREQENKYIELLAELISANISD 56
Qy 60 IDNFKRDKCALIKETVKQIRQIKQEKAAANIDEVOKSPVSTGGQVIDKDALGPM 119
Db 57 IDLSLVKPKDKILKKTYVDQIQLMKRMQEKSTDDQVOKSDISSSGVIEKESLGPL 116
Qy 120 LEALDGFEEVVALLEGVNVFVSENVQYLRVNOELMNKSVSILHVGHTEFEVKMLPKRS 179
Db 117 LEALDGFEEVVALLEGVNVFVSENVQYLRVNOELMNKSVSILHVGHTEFEVKMLPKRS 176
Qy 180 IVNGSGWSEGPERRNSHTFNCRMLVPLPDSEEGHNOEAHQKYTEMOCFAVSQPKSIR 239
Db 177 LVNGVPMPOEATRRNSHTFNCRMLVPLPDSEEGHNOEAHQKYTEMOCFAVSQPKSIR 233
Qy 240 EEGEDLOSLICIVARVYMKERPVLPSSSEFTTRDLOGKITSUDTSMRAAMKGMWDL 299
Db 234 EGEDLOSLICIVARVYMKERPVLPSSSEFTTRDLOGKITSUDTSMRAAMKGMWDL 291
Qy 300 VRCIOKFAHQEGESVYAKRHHHEVLROGLAFQIRFISLSDTLVAQKRSKILRSQ 359
Db 292 VRKCIYAFQ-POGREPREYARQLQEVMTRGASSPSYRFLINDTMSLAHTKCLCPQ 350
Qy 360 TTNEPOLYISLMLHREQNVCMNP-DLTGOTMKP-----LNPISNSPAHQALCSN- 412
Db 351 SPDMQPFIMGHIIDREHS--GLSPQDDTNGMSIPRYNPVNP--SISPAHVARSSTL 406
Qy 413 -PGQDMTSSNINFPINPKED-----KMPMGRCGGSGMHHVSGM-----QA 455
Db 407 PPSNSMNVSTRIN-----ROQSSDLHSSHSNNSNSGSCFCSGSOIVANVALNGQA 460
Qy 456 TTPQGSNTALKMNSPQSSQPGMNPQPTSMLSPRHRMSPGVAGSPRIPPOFSP-AGSLH 514
Db 461 SS-QSSNPSLNMNNSPMEGTGSLAQ---FMSPRQVYSGIATRMRMNNNSFPRIISTLS 516
Qy 515 SPVGYCSS--TGNSHYTNSSILNALQALSEHGVSIGSSLASPDLMKGNLQNSPVNMP 572
Db 517 SPVGYTSSACNNNNNSYINIPYTSLOGMNEGPNNNSVGFSASAPYLQKSSQNSPRLNIQ 576
Qy 573 PLKMKGISDSDCEFLVYEPSEGTGQAESSCHPEQKET-----NDPMLPRAVSSERADG 628
Db 577 P-AAESKDNKEIASILNEMI-----QSDNSSQDKPLDLSGLLHNNDL-----SDG 622
Qy 629 QSLHDSKQTKLQLLTTKSD-QMEPSPPLASSLD-----TKNDTGSJLPGSGSTHG 680
Db 623 DSKY--SOTSXKLVOLLTTTAEOQLRHADIDTSCADVLSCTGTSASANSAGSGSPSH 680
Qy 681 TSLKERRKILHRLLODSSPYDLAKLTAEATGKDISQSSSTAPG--SEVTIKQEPVSP 737
Db 681 SFLERHKLHRLLOE-GSPSDITTLVSEPPKKD-SASTSVSVTGOVGNSSIKLELDA 738
Qy 738 KKKK---NALLRYLLDKD-----TKDGLPEIPFKLERLSKTDPASTKLIAMKTEKE 789
Db 739 KKKESKHQRLRYLLDKREKDLRSTPNLSLDVKKYKVEKE-QMPCNTNPTPTKPTPE 797
Qy 790 EMSEFPDQPSSELDNLEIIDLON-SQLPOLFPDTPRGAPGVDAKOAIINDIMOLTA 848

```



Qy	681	TSLEKKHLLHRLLODSSSVYDLAKLTAEATGKDLSOESSNAPC---	SEVTKQEPSP	737
Db	681	SSLFAKKHLLHRLLOE-GSPSDITLTSVEDEKDD--SASRVSVTQOVQGNSSIKLELDA5	738	
Qy	738	KKKE---NALRLYLDKOD-----TKDIGPEPTPLKEFLDSTKTPASATKLIATKEKE	789	
Db	739	KKRSKSHOLLRLTKDDEKDLSTRNLSLDDYKVAKEKE-QMDPCNTNPTPTMKPPIPE	797	
Qy	790	EMSFEPGDQGSBELNDLEETLDDLQ--SOLPOLFPDTRGAPAGVSKOAILNDLMOLTA	848	
Db	798	EIKLEAQSPTADIDGEDQLLPTELEKAAQLPGICETDRMD---GAVTSVTKSEI----	849	
Qy	849	ENSVYTVVGAKOTALRISQSTFNNRPQGLRLLPQNPLDITTLQSPFGAGFPPIINS	908	
Db	850	-----LPASLQSAT-----	858	
Qy	909	SPYSVIRPQPGMNGOGMIGNLGNSGTGMGNASRPTMPSGEMAPQSSAVRYTCAAT	968	
Db	859	-----ARP-----	861	
Qy	969	TSANRPVQGMIRNPAAISIPMRSSQPGOROTLQSOVANIIGPELEMMGGPQYSQQA	1028	
Db	862	TSRLNR-----LP-----ELELEAIDNQGQGT	885	
Qy	1029	PPNOTAPWP-ESLIPIDQASFASQNRQPGSSPDLLCPHPAESPSDEGALLDOL--YL	1085	
Db	886	--GDQIPFTNNVTATNIQ---SKSEQICISSQDELCEPTTVEGRNDEKALLLEQVSFL	940	
Qy	1086	ALNRFOLEIEDBALGPELYSQAVID--PQFSSQDSN--IMLEOKAPVPQOYASQA	1141	
Db	941	SGKQETELALDRLALGIDKIVL-QGGGLDVLESEFPQQAATPLIIEERNLYSOPYSSE	999	
Qy	1142	QMAQGSYSPQADNFHTMGCRPSYATLRVQPRP-----GLRPTGLVN--QPNQL	1189	
Db	1000	PTAN-----LPSPFGVNRKPSLGTMPVQVTPRGAGSPQGMQPRQTLNRPAPANQL	1054	
Qy	1190	RLQLQHLQAO-----QNRPLANOISNVSNVNLTRPG---VPTQAPINQALQOR	1240	
Db	1055	RLQLQHLQAOQDLHNRQALINQFPAAPVGINRRSGMOQOITRPPPLNQMALQOR	1144	
Qy	1241	EILNQHLROKRMHQOQOVRTLMNGO--GLNMTPSVAPSGMPATNSNPRTIPANQO	1298	
Db	1115	ELVSQHQRLQI---QORALMLNQGSFGNLPSS---SGLPVQGNPLRPGADQO	1166	
Qy	1299	FPPPPNGISIQQDPDGGTG-----ATTPOSPL-----MSRNA	1331	
Db	1167	FPYPPNCTMPGTPPASTSFQSOLANPEASLANRSMYSRMTGNTIGQFQGTGINPQO	1220	
Qy	1332	HT-----OSPMQOSQANPAYQAPSDINGW	1356	
Db	1227	QNVQYQGAMVPOGEANFAPSLSPSSNVMPPIPPQSSLLQNPAPASQVSP--DKMW	1285	
Qy	1357	AQGMGMSNFQ--GSPPIFGQOANTSMYSNNKATINVSMAITNTGMSMOMTQOISMT	1414	
Db	1286	QOQAGIGNNVFSAOVANQP---TPAPPGVY-NMSTIVYSMAGANTNVOMNPMMAQOM--	1340	
Qy	1415	SVTSVTSGLSSGMPROVNDPALRGNDLPNLPDQPGMDMTKQEGDPTTK	1462	
Db	1341	--SSLQPMGNVCPEDINDPALRHGTLGICNQLSTDLKTEADGTQO	1386	
RESULT 11				
ID	P70366	PRELIMINARY:	PRT:	1405 AA.
AC	P70366			
DT	01-FEB-1997 (Tremblrel. 02, Created)			
DT	01-FEB-1997 (Tremblrel. 02, last sequence update)			
DT	01-DEC-2001 (Tremblrel. 19, last annotation update)			
DE	STERIOD RECEPTOR COACTIVATOR-1.			
GN	NC041 OR MSRC-1.			
OS	Mus musculus (Mouse).			
OC	Membralia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Eumetazoa; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			

OK NCBI\_TextId=10090;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=97008053; PubMed=8855229;  
 RX Yao T.P., Ku G., Zhou N., Scully R., Livingston D.M.:  
 RT "The nuclear hormone receptor coactivator SRC-1 is a specific target  
 of p300."  
 RL Proc. Natl. Acad. Sci. U.S.A. 93:10626-10631(1996).  
 DR EMBL: U64828; AAB38841.1; -  
 DR TRANSFAC: T04639; -  
 DR MGD: MGI:1276523; Ncoal.  
 DR InterPro: IPR001092; HLH\_dim.  
 DR InterPro: IPR000014; PAS.  
 DR Pfam: PF00989; PAS; 1.  
 DR SMART: SM00353; HLH; 1.  
 DR SMART: SM00091; PAS; 1.  
 KM Receptor.  
 SQ SEQUENCE 1405 AA; 152644 MW; 92088DAE1A6A7E77 CRC64;  
  
 Query Match 28.0%; Score 2137; DB 11; Length 1405;  
 Best Local Similarity 36.1%; Pred. No. 4, 9e-122;  
 Matches 578; Conservative 234; Mismatches 434; Indels 354; Gaps 65.  
  
 1 MSGMGENTSDPSRAETRRKRECP-DOLGSPKRNTEKRNREQENKYEELAEIFANFD 59  
 1 MSGLGSSSDPANPDHKKRGSPCDTLASS----TEKRRREQNKYLELAELLSANISD 56  
  
 60 IDNFNFKPKCALIKETVAQIQIKOEKAAANIDVQKSDVSSYSGGVIDKDALGPM 119  
 57 IDLSVKKPKCKILKKTVDQIQLMKRMEEKTTDDVDKSDISSSSGVIEKESIGPL 116  
  
 120 LEALDGFVVNVLEGNVNVSENVQYLRYNOBELMKNSVYSLIHGDTETVEYKNLPPS 179  
 117 LEALDGFVVNVCEGRIVYSENVTSYGLQNOBELMNTSVYSLIHGDAIEYKNLPPS 176  
  
 180 IYNGSWSGEPRRNSHTFNCRLVKKPLPDEEGHNDQEAHQKYETMOCFVNSPKSTK 239  
 177 IYNGVPMPEATRRNSHTFNCRLILHP--PEDPTGENEACORVEVMOCFTYSPKSTQ 233  
  
 240 EEGEDLDQGLIVARVVPKKERPVLPSSSFPTTRDLOKIKTSLDSTIRAKMPQEMDL 299  
 234 EEDGEFOQLCIARLP--RPPATIGVSEFMTKDPTGKIIISIDTSSLRAGRTGEMDL 291  
  
 300 VRCIOKFLPAHQEGEVSAKRRHHEVLQGLAFSQIYFESSDGTGLVAQKSKLIRSO 359  
 292 YKCIYAFV-QPQGRPSTAROLFQEVMTKGTASSSESYFLLNDGTMLSAHKKCICTYQ 350  
  
 360 TTNEPOLVLSLMLHREOVQYMPN-DLTGQTMGRP-LNPISN--SPAHQALSGN-- 412  
 351 SPDMQPFIMGHIIIDREHS--GLSPQDDNSGMSIDPRINP-SVNPGISPAHCVTRSTLP 407  
  
 413 PQQDMTLSSNIN---FPIGPKEQMGV--MGRFGSGGMHNVSGM----QATTPQGS 461  
 408 PENNNMVSAARVROOSSDLNSSSHTSSNNNGNCGCSGNDIVANVALNCOAGS-QSS 466  
  
 462 NVALKMNSQSSPGNNPQPTSMLSPRIRMSPGVAGSPRIRPQSP-AGSLSHVGVG 520  
 467 NPSLNLNNSPMEGTGLALSO---FMSPRQANSGLCTRAIRMNNSPPNIPPLTSSVGT 523  
  
 521 SS--TGNSHSYNSSLNALQALSEHGVSIGSSLASPDLMKGNLQASPNMNPPLSKMG 578  
 524 SCACNNNNSYINIPVTSIQGMNEGNNNSVGRSAGSPVLRAKQSSNSPRLSMOP-AKKE 582  
  
 579 SLDSKDCFGLYEPSEGTGQAESSCHGEQKET---NDPULPPAVSERADGOSRLHD 634  
 583 SKDSKEIASILNEMIQ--SDNSDANSNEKRLPDSGLLHNDRU-----SECSKY-- 630  
  
 635 SKGQTKLIQLLTQKSD-QNEPSPASSLSLT-----NKRSTGSLPGSG---STHGSLKE 685  
 631 SQTSHLVQLLTATTAEQQLRHADIDTSCDVLCTGTSSSASSNPBGCGCPSSH--SSLTE 689  
  
 686 KKKIILRLDDSSSPVDLAKLTAEATGKGLSOESSSTAAGS-----VTIKQPVSPKKE 741

```

Db 690 RHKLHLRLOE-GSPSDITTLTSLVEPEKDSVPASTAVSOGSGASIKLELDAARKE 748
Oy 742 ---NALLRYLLDKD-----TKDGLPEITPKLERLDSKTPASNTKLIAMKTEKESESF 793
Db 749 SKDHQRLRYLLDKREKDLRSTPNCLDVKYKVEKE-QMDPCNTNTPMTKRAPEEYKL 807
Oy 794 EPDGPSELNLEILLDDIOL-SQLPOLF-PDTRPGAPAGSVYKQAIINDLMQLTAEENS 851
Db 808 ESQGFADLDQFQQLPTLEKAQPSLCEFTBMDAVTQ-----VSIRAKE-- 854
Oy 852 PYTVVGAQKALRISQSTFNPRGQGRLLPNQNLPLDITLQSPGAGPPPIRNSAPY 911
Db 855 -VTPASLQPTTARA--PR--LSRL-----PELELAID 883
Oy 912 SVIRPQPMGNOGIGNOGNLNGSTGICNASRPTMPSGEMAPQSSAVRVCATTSYA 971
Db 884 NQEGQPG-----AGDQ-----IP--WANN--LTTIN 906
Oy 972 MNPVVGGMIRNPAASIPMRSSQPGQRQTLQSQVMNIGSELEMMNGGQYSSQQAAPPN 1031
Db 907 QNRK-----EDQGISQL----- 919
Oy 1032 QTAPWESILPTDQASASQNRQFGSSPDILCPHAAESPDEGALLDQ--YLALRN 1089
Db 920 -----DELCPPTVEGRNDEKALLEQVSELSGKD 950
Oy 1090 FDGLEIDRALGPELYSQAQVD--PEOFSQDSN--IMLEOKAPFPOQYASQAQMAQ 1145
Db 951 ETELAELDRALGIDKLY-QGGGLVYLSERPPQATPPLIMEDRPTLYSQYSSPSPLA- 1008
Oy 1146 GSYSPODPNHTMGORBSYATLMQ-----PRGLRPTGLYON--QPNQLRLQL 1193
Db 1009 GLSGPFG-----AVRKPSLGMFVQYTPRGTFSPNMGMQPROTLNRPAPAPQLRLQL 1064
Oy 1194 QHRLOAQ-----QNRQPLANOISVNVNLTLPQ-----VTOAPINQMLAQORELIN 1244
Db 1065 QQRLOGQOQLMHNQKAILNQFANAPAGNMRSQMGOQITPOPPLNAQMLAQORELYS 1124
Oy 1245 QHLRORHQQOQVQOATFLMARGO--GLNMTSPVAVSGMPTATMSPRIIPANQOQPEP 1302
Db 1125 QQHQRQRIQ-----QQRAMLMRHSFGNNIPPS-----SGLFVQMGTRPLRPGAQOQFFYP 1176
Oy 1303 PNYG-----IS 1308
Db 1177 PNYGTNGTPRPASTSPFSQLANPEASLATRSSMVNRCMAGMNGQFGAGISPMQOVVF 1236
Oy 1309 QQPPRG-----FTGATTPQSPILMSPRMHNQSPMMQOQOANPAYQAPBDINGMAQGN 1360
Db 1237 QYPPGLVPGQEAFTABLSFGSSMVPVPYPPQSSLLQOTPTPSYQSP--DMKAWQGT 1295
Oy 1361 MGSMSFQ--QSPPHFGQOQANTSMYSNNNININYSMATNTGMSMSMONTQOISMTSVTS 1418
Db 1296 MGNNVFSQAQVQSP--APRQPGVY--NMSITVSMAGNANNINQNMNMQMOM-----SS 1348
Oy 1419 VSTGLSGMPEQVYNDPALGCGNLFPPNOLPGMDMIKQEGD 1458
Db 1349 LQMPGMNTVCSQMDNDPALRHTGTGLYCNQLSSTDLTKTDAD 1388

```

```

RN [1]
SEQUENCE FROM N.A.
RA Kamei Y., Xu L., Heinzl T., Torchia J., Kurokawa R., Gloss B.,
RA Lin S.C., Heyman R.A., Rose D.W., Glass C.K., Rosenfeld M.G.;
RT "A csp integrator complex mediates transcriptional activation and ap-1
RT inhibition by nuclear receptors.";
RL Cell 0:0-0(1966).
DR EMBL; U56920; AAB01228.1; -.
DR MGD; MGI:1276523; Ncoal.
DR InterPro; IPR001092; HLH_dlm.
DR InterPro; IPR000014; PAS.
DR Pfam; PF00989; PAS; 1.
DR SMART; SM00353; HLH; 1.
DR SMART; SM00091; PAS; 1.
DR Receptor.
SQ SEQUENCE 1405 AA; 152572 MW; BD50F8FD62D852A CRC64;

Query Match 27.6%; Score 2108; DB 11; Length 1405;
Best Local Similarity 35.7%; Pred. No. 2,9e-120;
Matches 572; Conservative 236; Mismatches 434; Indels 362; Gaps 65;

Oy 1 MSGGENTSPSPRAETRRKECP--DOLGSPKRTTEKRNDEKRYTEELAEIIFANFD 59
Db 1 MSGIGDSSDPANPDSHKRGSPCDTLASS-----TEKRREQEKYLEGLAELLSANISD 56
Oy 60 IDNFEKPRDKALIKETVKQIRQIKQEKAAANIDEVOKSDVSTSGQYIDKDALGPM 119
Db 57 IDLSLTVFPMCKILTKITVDQIQLMKRMQEKSTDDVOGSDSIDSSQGVIEKESLPL 116
Oy 120 LEALDGEFFVNVNEGAVNVSENVTOYLRNOELMKNVSYSLHVGDTHEFKNLLPKS 179
Db 117 LEALDGEFFVNVNEGAVNVSENVTOYLRNOELMKNVSYSLHVGDTHEFKNLLPKS 176
Oy 180 IVNGSGWGPERRNSHTFCRMLVKPLPDEEGHNOEAKHQETMOCFAVSQPSK 239
Db 177 LVNVPWPQEAETRNSTFCRMLIHP--PEDGTNOEQCYEVMOCTVYSQPSK 233
Oy 240 EEEGDLQSLICVARYRPMKERPVLPSESEFTTRQDLOGKITSLDSTMRAMKPGMED 299
Db 234 EDGEDPQSLICLARLP--RPAITGVESFMTQDTGKTIISDTSLSAAGTGMED 291
Oy 300 VRRCIQFHAHQEGESVYAKRHHEVLRQGLAFSQIYRSLSDGTLYAAQTSKILRSQ 359
Db 292 VRKCIYAF--QPGREPSYAROLFQEVMTRGTAESPYSRFLNDGTMLSAHTCKKLCY 350
Oy 360 TTNEPOLVISLMLHREONCVNMP--DLTGQTMKP--LNPSSN--SPAQALQSGN-- 412
Db 351 SPDQPFIMGIHITDRHS--GLSPQDSSNGMSIPRINP--SVNPGISPAHGVTRSSLP 407
Oy 413 PGQDMTSSNIN--PPIGPKBQMGMP--MGRFGSGGGMNV-----SGMQAT 457
Db 408 PSNNMVSARVNRQSSDLNSSSHTNSNNQGFQSPGNOIYANVALNOGQAGSQTN 467
Oy 458 PQGSNYALKNMSPQSGPMNPGOPTKMLSPRIHMSGCVAGSPRIIPSOQSP--AGSLHP 516
Db 468 P-----SLNLTNNSMEGTGIALSQ--FMSPRRANGGLTRAMSNNSPPNPITLSP 519
Oy 517 VGVCS--TGNSHYTNSLNAQLSEGHVSLGSSLASPLDKMGLNONSPPVMMNPPL 574
Db 520 VGIISGACNNNNRKYSNIPYISLQGANEGPNNSVGFAGSFVLRQMSQNSPSRLSQP- 578
Oy 575 SKMGLSKDCDFGLYGPSEGTGQAESCHPQOKET---NDPNLPVPAVSSERADQGS 630
Db 579 AKASXSKSKEIASILNMIO--SDNSQNSANEKGPLDSGLLHNNDRL-----SEGD 628
Oy 631 RLHDSKQTKILLTLTKSD-QMEPSPLASLST-----NKDSTGSLPQSG--STRGT 681
Db 629 KY--SQTSXKLVQLLTTAEQQLRHADIDTSCKDVLSCTGTSSASNSPSCGTPSSH-S 685
Oy 682 SLKEKHILHRLLODSSVDLAKITFAEANGKDLQSSSSAPASE-----VTIKQEVSP 737
Db 686 SLTERKHLRLOE-GSPSDITTLTSLVEPEKDSVPASTAVSOGSGASIKLELDDAA 744

```

RESULT 12  
 ID 061202 PRELIMINARY; PRT; 1405 AA.  
 AC 061202;  
 DT 01-NOV-1996 (Tremblrel. 01, Created)  
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 DE STERIOD RECEPTOR COACTIVATOR 1A.  
 GN NCOAL OR SRC1A.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;



[illegible][illegible]



OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Raineri I., Soler M., Senn H.,  
 RT "Analysis of Human Immunodeficiency Virus type 1 promoter insertion in  
 RT vivo."  
 RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; U19179; AAA64187.1;  
 SO SEQUENCE 416 AA; 45239 MW; 68AF955508339CAD CRC64;

Query Match 7.1%; Score 545; DB 4; Length 416;  
 Best Local Similarity 33.6%; Pred. No. 2.1e-25;  
 Matches 142; Conservative 63; Mismatches 110; Indels 108; Gaps 16;

QY 1126 LEKAPVFPQGVASQAMQAGSYPMDPFHMGQREPSATLRMPRP-----GL 1176  
 DB 1 MEERPNLYSQPSSPSPTAN-----LPSPQGVNRQKPSLGTMPVOYTPRGAFSPQMG 55  
 QY 1177 RPTGLVON--OPNQLRLQOHRLQAO-----ONROPLMNQISNVNLTLPRG---V 1224  
 DB 56 QPQOTLNRPRAARNQLRLQRLQGLQGLQHONRAIILNQPAATAPVGINMMSGMOQOI 115  
 QY 1225 PTQAPINQMLAQRREILNQHRLQRQMHOQGVQOFTLMRGO--GLNMTPSMVAFSGM 1282  
 DB 116 TPQPLNAQMLAQRREILNQHRLQRQMHOQGVQOFTLMRGO--GLNMTPSMVAFSGM 167  
 QY 1283 PAMNSNRITQANAAQPPFPNPGISQPPRGFTG-----ATTPQSL 1325  
 DB 168 PVQMGNRLPQAGAPQPPYPPNPGTPTPASTSPESQLAANEASLANRNSMVSFGMT 227  
 QY 1326 -----MSPRMAHT-----QSPMAQ 1340  
 DB 228 GNIGGFGTGTINQMOQNVQYPCAGVPCGEANFAPSLSPSSMVPMPITPPQSSILQ 267  
 QY 1341 SQANPAYQAPSDINGMAQMGNSMFSQ--QSPPHFGQANTSMTSNNNINVSATNT 1398  
 DB 288 TPASGYSQSP--DKMAQQAIGNNVFSQAVQNP--TPAQGVY--NNMSITVSMAGN 342  
 QY 1399 GGSMSNMQMTGISMSTSVTSVTSGLSSMGPEQVNDPALGSLNFPQGLGMIMIKQEGD 1458  
 DB 343 TYNQNNPMAQOMOM--SSLOMPGAMNTVCPQINDPALNHTGLYCNHLSITDLKTEAD 399  
 QY 1459 TTR 1461  
 DB 400 GTQ 402

RESULT 17  
 Q9BR49 PRELIMINARY; PRT; 203 AA.  
 AC Q9BR49;  
 DT 01-JUN-2001 (Tremblrel. 17, Created)  
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)  
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 DE D11049G16.2.2 (CONTINUED FROM BA456N23.2 IN EM:AL353777 AND DJ237J2.1  
 DE IN EM:AL021394) (FRAGMENT).  
 GN NCOA3.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Skuce C.,  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL034418; CAC36067.1;  
 DR InterPro; IPR000014; PAS.  
 DR SMART; SM00091; PAS; 1.  
 FT NON TER 1  
 SO SEQUENCE 203 AA; 23303 MW; 895D62D858C3B04C CRC64;

Query Match 6.7%; Score 508; DB 4; Length 203;  
 Best Local Similarity 56.7%; Pred. No. 1.4e-23;  
 Matches 102; Conservative 24; Mismatches 44; Indels 10; Gaps 2;

QY 122 ALDGFVVNLEGNVYFVSENVQYLRYNQDELANKSVYSLHVGDTFVKNLKPSIV 181  
 DB 1 ALDGFVVNRDGNIVFSENVQYLRYNQDELANKSVYSLHVGDTFVKNLKPSIV 59  
 QY 182 NGVSWGEPERRSRSHFENCMLYKPLPDSSEEGHNDQEAHOKETQCFVAVSPKSKKE 241  
 DB 60 NGVSWTNEQROKSHTFNCRMLKTPHDILEDINASPEKROKETQCFVAVSPKSKKE 119  
 QY 242 GEDLOGLICVARNVPMKEPVLPSSEFTTRDQGLKTSITDSTMAAMKPGWEDLVR 301  
 DB 120 GEDLOGLICVARNVPMKEPVLPSSEFTTRDQGLKTSITDSTMAAMKPGWEDLVR 170

RESULT 18  
 Q9GS19 PRELIMINARY; PRT; 2035 AA.  
 AC Q9GS19;  
 DT 01-MAR-2001 (Tremblrel. 16, Created)  
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)  
 DE BCDYSONE RECEPTOR CO-ACTIVATOR TAIWAN.  
 GN TAI OR CG13109 OR CG18494.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Bai J., Uehara Y., Montell D.J.,  
 RT "Regulation of Cachectin-Mediated Cell Motility by Taiman, a Drosophila  
 RT Protein Related to Atb1, a Steroid Receptor Coactivator Amplified in  
 RT Breast Cancer."  
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY008258; AAG1637.1;  
 DR FlyBase; FBgn0041092; tai.  
 DR InterPro; IPR001092; HLH\_dlm.  
 DR InterPro; IPR00014; PAS.  
 DR SMART; SM00353; HLH; 1.  
 DR SMART; SM00091; PAS; 2.  
 KW PROSITE; PS00583; PKB\_KINASES\_1; UNKNOWN\_1.  
 KW Receptor.  
 SO SEQUENCE 2035 AA; 213790 MW; 6EC61F9E8447B1A5 CRC64;

Query Match 6.5%; Score 495.5; DB 5; Length 2035;  
 Best Local Similarity 19.9%; Pred. No. 2.2e-21;  
 Matches 365; Conservative 218; Mismatches 623; Indels 627; Gaps 79;

QY 1 MSGMGENTSDPSAETRRKRECPDGLGSPD--KRNTKRRNRQENKYEELAEIFANFN 58  
 DB 209 ISANSANSKPSASGR-KIRBRKDSKVNLPOQINCKNNEKRREARENGYIEQLSEILFNKR 267  
 QY 59 DIINFENKPKCAILKEFYQIQI--KEQEKAAA----- 92  
 DB 268 G-DMTSTPKDAKAILQVYTYREICDKGNRDISSTNNNSTTTNNNTNSNNNNNT 326  
 QY 93 -----NIDVQKSPVSTGGGVIDKD--ALGPM-----MLELDFG-- 126  
 DB 327 SKQATSTKCSRCAITDNCSTHPVQGGVSTPEPDLPEPSILLQVETASVYFALPHYIS 386  
 QY 127 -----FFVNLLEGVYFVSENVQYLRYNQDELANKSVYSLHVGDTFVKNLKPSIV 181  
 DB 387 GVGWVLLQVWANGIIESCQINRDLIGYERQELYHQPLVWYLSGSHAKL-----EPIT 440  
 QY 182 N-----GGSWS 187









```

QY 1031 NQAPMPESLIPIDQASFAQ-NRPPGSSPDDLCPHPAESPDGALLDLYALRN 1089
DB 1138 --TSP--VVSTTSGDMTQSGSTQIPGSTGTVTQPTSGTSGSTSGELTISQ----- 1184
QY 1090 FDLLEIDR-ALGIPELVSSQ-----AVDPEQ 1116
DB 1185 --GSTQPPKSLSTSPAIISTSTQOVSSTNPGSTVTQPTSTVGRSGTSGTGTGSS 1242
QY 1117 FSSQDSNIMLEQAPVFPQOYASQAQMAQSSVSPMODPN-----FRTMGORPSVATLR 1169
DB 1243 TSSSSSATSLSSSSPV-PTSTQSPNPSTSGSTPTTPSPQSTSPVYSTTGEMTSHGSTQ 1301
QY 1170 MQRPGIRPGLVQNPQRLQRLQRLQAQNRKQPLMNOISVSNVNLTLRPG----- 1223
DB 1302 TPSTIGSTVY-----OPSTVSGNSGSTVTIGSSEASTSG 1337
QY 1224 -----VPTQAPIAQMALQORQREILNHLRQRMHQOQOQQTLMNRQGLN 1271
DB 1338 SSFKTSPSSISPYTSSPITSTFAS-----STSGSTIS 1371
QY 1272 MTPSNVAPSGMPATMSNPRIPOANAQF-----PPPNYGISQDPDPTGA 1318
DB 1372 DVSSVETTLAPLSSSLPSTVPSSTQSFSSSTSGSSKASSPVPSQTSPTPTMGSTES 1431
QY 1319 TTPQSPMLSPMAHT-----QSPMMQOQANPAYQAPSDING----- 1355
DB 1432 STLSSTISGSGTHTTMSKASSGSTSPSTNSQTSTVTMGSSSTGSTVSSASTQPOMS 1491
QY 1356 WAGQNNNGNSM-----FSQOSPFPHQOANTSMYNNMNTNVSATNTGQMSSNMQ 1407
DB 1492 TSGSSSNGSTVASTSPASSSTAPSSSTGTMSTST--SGTVGSTISSTTASASQOTGS 1549
QY 1408 TQISMTSVTSVSTGSLSMGPEQVNDPALRGNLFPNQLPGM 1450
DB 1550 TYVMGSSSTSGVSTSSASTSQPMSTSQGSSASSTVASTAGL 1592

RESULT 22
Q9N693
ID Q9N693 PRELIMINARY; PRT; 2302 AA.
AC Q9N693;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, last sequence update)
DE 01-JUN-2001 (TREMblrel. 17, last annotation update)
DE BRKLEBS-B (SCRIBLER LONG ISOFORM).
GN SBB BR BKS OR CG5580.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydriidea; Drosophilidae; Drosophila.
OC NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20265906; PubMed=10804172;
RA Yang P., Shaver S.A., Hilliker A.J., Sokolowski M.B.;
RT "Abnormal turning behavior in Drosophila larvae: identification and
RT molecular analysis of scribler (sbb).";
RL Genetics 155:1161-1174(2000).
DR EMBL: AF242194; AAF70256.1; -.
DR FLYBASE: FBgn0010575; sbb.
DR InterPro: IPR003880; Phosphopant_attach.
DR SMART: SM00355; Znf_C2H2.1
DR PROSITE: PS00012; PHOSPHOPANTETHEINE; UNKNOWN_1.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.

```

```

DR PROSITE: PS00157; ZINC_FINGER_C2H2_2; 1.
KW DNA-binding; Zinc-finger.
SQ SEQUENCE 2302 AA; 230013 MW; B35F759AE2C21729 CRC64;

Query Match 4.5%; Score 340; DB 5; Length 2302;
Best Local Similarity 20.5%; Pred. No. 8.5e-12;
Matches 280; Conservative 148; Mismatches 481; Indels 460; Gaps 62;

QY 369 SLMLHREQVNCVNMNDLDTGOTMGKPLNPISSNSPA--HQALSGNPGQMDLTSSNINF- 425
DB 968 STFSHAQVHAKLRN-----GATKRGKATRSAGMAAANSSSSSGMGGATPSTPTAFL 1023
QY 426 -----PINGREQMGMPMGRFGSSGMNHS-----GMQATTPOG 460
DB 1024 PPRPEKRKSDEAPSPPLNGASD-GASVGGIGGAGGVNMVNASGIPISASGGGLATQPOPS 1082
QY 461 -----SNVALKNSPSCSSFGM--NGQPTSM-----LSFRHMSPG 495
DB 1083 LLNPVTGLNVQISTKCKKTASPCALISPVLLCEPEODCSKKYKHANGLRYHQSHAHGAGG 1142
QY 496 VAG-----SPRTPSQFPAGSLHSPVGCSTGSHSTYSNALQALSGHG 545
DB 1143 ASSMEDSMQAPDPDPAITPSP-----GVASGTSGASVASSAVPA-TAPSAGQ- 1189
QY 546 VSLGSSLASPELKMGNLONSPVMN-----PPPL 574
DB 1190 ---GTYAVSPNTPLANSSNVTNGVNAFSAFATGVTIAAPNTTPTVETQAPLTGPPEY 1246
QY 575 SKMGSLSKDCFLGYEPSEGTTCQAESSCHPEQKETND---PMLPRAVSSERADGOS 630
DB 1247 T-----PPAPFPICAVAT--PCAEQSVSSVPLPLGNLPLTAGNSATQOQ 1288
QY 631 RLHDSKGQFTLL---QLLTTKSDQMEPSPLA-----SSLSDTKNDSTGSLPGSS 676
DB 1289 QPPTQOQPOLVPPGSAASLQOQOQOQOQVAGGSTITAGISGALSHQOQMLGGLP--- 1345
QY 677 STHGSLKEKHKILHRLDQSS-----SPVD---LAKLTAEATGKDLQSSSSTARQ 725
DB 1346 ---AMLSDDQO--QALLQOQALKAGVLRGPDPGNPLQOQOQPAQAVNPQTQSPRP- 1398
QY 726 SEVTTIQEP-----VSPKKEN-----ALLRLDKDQTKDI 757
DB 1399 SHVQDDQTPSAVAQAQGLTSPGFSGVGAASAKKQKNNKSPGDPFEGRSREDOVSP 1458
QY 758 GLPEI---TP--KLRLDSKTDPAANTK---LIANKTEKEESFPDQPGSELNLEE 808
DB 1459 AYSDISDDSTPVAEQEMLDKSVGAQAVTAKHTELMKKPTEVGVGVPFPPAPMNYVGMTQ 1518
QY 809 ILDDLQNS-----QLPOLFPDTRPGAPAGSVDRQAIINDLMQLTENSPLYT----- 854
DB 1519 FYPAQOQASAPPQOQOQOQOQYVWQTEBQKPPGLPALVQAQOQOQOQPGAPPTSQPPSH 1578
QY 855 ---PVAQQTALRISQSTFN-----PRG-----QUGRLPNQNLPLDI- 891
DB 1579 LLGPPQOQSVAAHLADYSGKNKDPDLMTKPPQPGQPPSQOQOQSGOLSGDENNGKDVG 1638
QY 892 TLQSPGAGFPPIRNS-----PYSVIPOGM-----MGNQGMIG----- 927
DB 1639 PPTSQPGSQP-PPVNIASAAGPPPGSLP-PGLGGLSALGAAGIGCGPKGMHFTFPNF 1696
QY 928 -----NQNLSNSTGMIGNSASRPTWPSPGEMAPQSSAVRVTCAATTSAMNRPV 976
DB 1697 IPPAVYVNDPNNGSVSIYASEAAKLSGHGLPSPSSQAQQLSGISIK----- 1744
QY 977 QGGMATNPAASTIPRPSQPGQROTLOQOVMNIGP--SELDMMNGGQYQSOQAAPPQQTAP 1035
DB 1745 EERLKESP--SPHDQPKHPSQOQMTASKLIKQEPMTKQEIKEPNSNPGQGHPPQOQP 1802
QY 1036 WPESLIPIDQASFAQNRPPGSSPDDLCPHPAASPSDEGALLDQVLALRNPGLEE 1095
DB 1803 APO---PQOQOQPPPPQOQPHALHPKDL----- 1827

```







QY 1390 INVSMATNGMGSSNMOTGOISMTSV-----TSVSTISGLSSMGPEQVNDPALRGNIFFPN 1445  
 Db 4059 KD---STPPGHVPTPTPMAKTSKTSSESVIRPTPTGLAVISANTVGS-LTREENLIKI 4114  
 QY 1446 QLPGM-MIKOE 1456  
 Db 4115 SOPKDELEED 4126

RESULT 26  
 09VPL2 PRELIMINARY; PRT: 5533 AA.

AC 09VPL2; 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE SPEN PROTEIN  
 GN SPEN OR CG18497.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_Taxid:7227;  
 RN 111  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BERKELEY;  
 RX MEDLINE-20196006; PubMed-10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,  
 RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K.J., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,  
 RA Glodde A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harits N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Housh K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jaitani M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Mlshina N.V., Mobarry C., Morris J., Moshirel A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclib J.M.,  
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Stronig M., Sun E.,  
 RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Maassarman D.A., Weinstock G.M., Weisenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AE003590; AAF51534.2; .  
 DR HSSP: P09651; IHAL.  
 DR FLYBASE: FBgn0016977; spen.  
 DR InterPro: IPR000504; RRM.  
 DR Pfam: PF00076; rrm. 3.  
 DR SMART: SM00360; RRM. 3.  
 DR PROSITE: PS50102; RRM. 3.

SQ SEQUENCE 5533 AA; 597145 MW; 713E21117EAF2CCD CRC64;  
 Query Match 4.4%; Score 334; DB 5; Length 5533;  
 Best Local Similarity 20.3%; Pred. No. 7e-11;  
 Matches 331; Conservative 197; Mismatches 654; Indels 450; Gaps 74;

QY 8 TSDPSRAETRRKREKCPDOLGSPSKRNTKRRNRBOENKYEIELAELIFANPDIDNENFKP 67  
 Db 2819 TSAPSTRAQTSKRG-----EDKKMEFIIGIISDEESQF-P 2851

QY 68 DKCAILKETVKKQIRKQEKAAANIDEVQKSDVSTGCGVIDKALGPMLEALDGF 127  
 Db 2852 EQA-----ETNK-----DIIPSSVST-----GPVSAALQ--- 2877

QY 128 FVYNLEGNAVVESENVQYLRYNOEELMKNYSILHV---GDHTEFYKNLIPKSIYNGS 185  
 Db 2878 -----TYKQEPSTPKSKMEAHILQTLVHEPEQOOLERSRLSGGS 2917

QY 186 WSGEPPRRNSHTNCRMLVPLPDSEEGHNOEHOKEVETMOCAFVSQPKSI-KEEGE 243  
 Db 2918 SSSSHADREHRRREKR--EKRRKESQREQNQ--HQKSKVET-KVDDONSVDMEAGR 2973

QY 244 DLOSLCIVARVPKMERVLPSSSEFTTRQDLQKTTSLDTSTMAAKPGMEDLYRRC 303  
 Db 2974 ALQAQL-----MSDFDKPISE-----EATPSTATYRSDMTD----- 3006

QY 304 IQFHAQHEGESVYAKRHNEEVLROGLAFSQYRFSLSGTLVAQTKSLIRSQTTNE 363  
 Db 3007 VRFPSNEDNNSV-----DMTKQCVKSEQDQHKSKD-----KKKKRRSKREKO 3051

QY 364 POLVLSHMLHRBQNVCMNPDLTGOTMGK-----PLNPSSNSPAHQALC 409  
 Db 3052 EKL-----LQQRRLSRPVASTSAPPLPKGLTVNVQAASKHADLDQDAKHISPP--PVC 3105

QY 410 SGNPQDMTSSNINPFINPKQMGPMKPRFGSGGMNVSGMOTTPGSGVYALKMS 469  
 Db 3106 KPSPSLPCLIGDDDDALHTPKAKPTTPSSR--GNGL-----T 3142

QY 470 PSQSPGMNPGQPTSMILPBRHMSPGVAGSPRIIPSPFSPAGLSHPVGSSGTNSHY 529  
 Db 3143 PSREKP-----RLSP-----IPKPTIANSSTLSTQAEIYV-----SSGVIYS- 3182

QY 530 TNSLNAALQALSGHGVSLGSSLASPDLMKGNLQNSPVNNPPLSKMGSLSKDF--G 567  
 Db 3183 -SSALATTPPTSPYAAAGS-----AAP-----GLDNSPT-----SASACKKKESEFIPG 3224

QY 588 LVGEPSGCTGQAESSHCHPEQKET-----NDPNLPRAVSERRADQGRSLDSKQTK 640  
 Db 3225 FDSQLDRLISESAVQST--SAEFNSTSLDNIADPEKIPVAPPRATKPLDKLEESKSRVT 3283

QY 641 LLQLLTTSQDMPEPPLASSLSDTNKDSGSLPGSGSTHSTLKEKHKILHRLDSSSP 700  
 Db 3284 ISQ-----EEFESAVNALGSPFTSSTTDVSLDGMEMSVNLEFPTL---VIAEDEE 3336

QY 701 VDLAKLTAEATQKGLS-----QESSSTAPGSEVYTIKQEPV---SPKKEMALLRY 747  
 Db 3337 AALAAAIETAGCPASTLEPEMEPEREAPDDPEAKEISEPVEVLEDEELKAV--Q 3394

QY 748 LLDKDKTKDGLDELTPKLER---LDSKDPANTKLIAMKTEKEKMSFERGQOPSELD 804  
 Db 3395 SLKHEDMDIKAD--TPQSERDQIDITVEENDEADSSGSLKIDETVQSSSSPEKSIS 3452

QY 805 NLEELIIDDQNSQDLPOLPEPTRGAPAGSVDKOAILNDMLQALAENSPPV--PGAQKTA 863  
 Db 3453 NNSPTREPRANIDIPV--ESQKLKSNESIPQSVITTKLPFL---DTPKTPAGLPPSPV 3507

QY 864 RISQSTFNN-----PRPOLGR-LLPNQLPLD----- 890  
 Db 3508 KIEPPTISKQGPLVQPVQTVLPAHPSTSGISANSVINLDSNVISCSNTSASATAS 3567

QY 891 ----ITLQSTGAGPPPIRNSSPYVITPQPGMMGNGMIGNGNLGNSSTGMIIGNSASR 946  
 Db 3567



```

Db 3568 ASASISFGSPAS-----QANMPDASTPKOGRITPQOAIKRTOSLIMOPTISIEQTPH 3621
Qy 947 PTMSEGAPOSASAVRYTCATTSAMNRPVOGGM-----IRNPASIPMRSSQPGOROTL 1002
Db 3622 FAVPMVLSPS-----HHPOGRYMGIRAPDS---PSPHSPKGVAAQ 3664
Qy 1003 OSOYVNIPTSELMNMGSPQYS-----OOQAPPNQ-----TAPWESILPIDOAS--VAS 1050
Db 3665 SRVGOQLSP-----VGRPMVSQSPQOQVOOTQOQHALLITSPOSNTSPLASPTTVLS 3718
Qy 1051 QNRPPFS-----SPDILCP-HPAESPDEGALLDQYALRNFDLEETDRALGPE 1104
Db 3719 SSNSPTTSKVSQYPRNQOQVQOQSPKSVAEVOTTPQMTIPLQMTPIQYVPH-----PT 3774
Qy 1105 LVSOQADRPQFSSQD-----SNIML-----EQKAVYFQOYASQAMQAG 1146
Db 3775 IISKVYVTPQOATQOQVASSPRLGSLPRHKNVHLNANOQOQPOVIAKMTAHOHQHMQ 3834
Qy 1147 SY-----SPMDPFTMGORPSYATLIMQPRPLRPTGLVQNOP-----NOLRL--- 1191
Db 3835 QFMHQMTQROQHMQOQOOLHGOQOITSA-----FOHQHMQHQAQOQOQHNNQOHLMQ 3889
Qy 1192 -----QLOHRLQAQOQNRQPLMNOISVSNVNLTLRPGVTPQAPINAMLAORQREIL 1243
Db 3890 LHAQOHPKOKHOAQOQCFNQOIQOHQSOQ-----QHVOQOQOQOQHLSQ 3935
Qy 1244 NQHLRQRMHQOQOYQOQTL-----MMRG-QGLMTPSMVABSG-----MPATMSPR 1290
Db 3936 QOHOSQOOLNHOHQOQOOLQOIKLOQMHGPOQOQKSPQVGHGIGSTSIFFASQOHSQ 3995
Qy 1291 IP-----QANAQOF-----PPRPYGIS-----OQDPGFTGATTPQSPILMSPRMAHTOS 1335
Db 3996 LPAQVPOQOHPQOLSHSPCKPMTLVSVNGQVORPALITRVGHSIQOQOOLPHQOSS 4055
Qy 1336 ---PMPQOQO---ANPAYQADSDINGMAQNGMNGSMFSGSPHFGQOQANTSMYSNMN 1389
Db 4056 SGHHPQKOLSSPGANLPLQOTPLNVLOMTPKILVQOHIVAQOVPPPOQNAIHYPQNOG 4115
Qy 1390 INVSMATPTGMSMNQMTGQISMTSV-----TSVTSGLSSMGPRQVNDPALRGNLFPN 1445
Db 4116 KD---STPPGHVEPPRPAASAKTSBSVSVIRPTPTTGLAVISANTVGS-LITEENLIKI 4171
Qy 1446 QLPQMD-MIKOE 1456
Db 4172 SQPKODELIED 4183

```

## RESULT 27

```

Q9NHN1
ID Q9NHN1 PRELIMINARY; PRT: 5554 AA.
AC Q9NHN1;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE SPLIT ENDS LONG ISOFORM.
GN SPEN OR CG18497.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Kang B., Wu S., Shin Y.-A., Luo L., Kolodziej P.;
RT "Split ends encodes large nuclear proteins that regulate neuronal cell
RL fate and axon extension in the Drosophila embryo."
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF221715; AAF34661.1;
DR HSP: P09651.1HAL;
DR FlyBase; FBgn0016977; spen.
DR InterPro; IPR000504; RRM.
DR SMART; SM00360; RRM; 3.
DR PROSITE; PS50102; RRM; 3.

```

```

SQ SEQUENCE 5554 AA; 59918 MW; 4037E27833D0C622 CRC64;
Query Match 4.4%; Score 334; DB 5; Length 5554;
Best Local Similarity 20.3%; Pred. No. 7.1e-11;
Matches 331; Conservative 197; Mismatches 654; Indels 450; Gaps 74;
Qy 8 TSDPSRAETRRKKECPDGLSPKRNTEKRRNREQNKYTEELAEILFANFNIDIDFNKRP 67
Db 2813 TSAPSTQTSKRG-----EDMEFTIGIISDEESQF-P 2845
Qy 68 DKCALKEFTVQRIQIKOEKAANAIDEVOKSDVSYSGQGVIDKADAPMLLEALDFF 127
Db 2846 EQA-----ETNK-----DILPSVST-----GPIVSALQ--- 2871
Qy 128 FVNINLEGNVFESENVOYLRYNOELMKNKSYSLIHW--GDHFEVKNLLPKSLVNGS 185
Db 2872 -----TYKQESTPNKNEEAHQLIVHEPEQOQOQLERSRLSGS 2911.
Qy 186 MSGERPRRNSHTFNCRMLVYKPLPDEEEDHNOQEAHQYETMQCPAVSQPSI--KEEG 243
Db 2912 SSSSHADERRHRRERK--EKRRKESQREQOQ--THQSKSVET--KVDNDSVDMDEAGR 2967
Qy 244 DLOSCLICVARRVPKERVLPSSSEFTTODLQKITSLDTSTYRAAKRGMEDLVARC 303
Db 2968 ALENQ-----MSDDTPRISF-----EATPSTAAVYRSDMTD----- 3000
Qy 304 IQKFAHQEGESVYAKRHHHEVLRQGLAFSQIYRESLSDGTLYVAQTKSLIKLSQTTNE 363
Db 3001 VEFESDNDNNSV-----DMTKQVSEKQEQEHKSKD-----KKKKKKRKEEKQ 3045
Qy 364 POLVLSLMLHREQNVCMNPDLTGQTMK-----PLNFISSNPAHQALC 409
Db 3046 EKL---LQOQRESLPNVAISSAPPTPGKLTVNVQAASKHADQLDKKHITSP--PVC 3099
Qy 410 SGNPGQDMLTSSNINFPINGPKQMGMPGFRFGSGGNHVSQMATTPQGSNTYALKMS 469
Db 3100 KPSPLPCILIDDDDLAHTKAKPTTPSSR--GNDGL-----T 3136
Qy 470 PSQSPGKMPQOQPTSMLSPRHRMSPGAVGSPRIQSPQSPAGSLSPVGVCSSTGNSISY 529
Db 3137 PSREKP-----RLIEP-----IPKPTIANSSTLQSAETPV-----SSGTVIS 3176
Qy 530 TNSSLNALQALSEGHVLSGSSIASPDLMKGLQNSPVNMPPLSKNGSLDSKCF--G 587
Db 3177 -SSALATTPSTTAGVS-----AAP-----GLDNSTP-----SASAQCKKKSFTPG 3218
Qy 588 LYGEPESECTGQAESCHPGEQKET-----NDENLPAYVSEERADQSRHLDSKGQTK 640
Db 3219 FDGQLDDRISESAGVSI-SAEFNSISLDNTADEKTIIVASPPRATKPLDKLEBSKAVT 3277
Qy 641 LIQLITTSQDMESPPLASSLSDTNKSDTGSLSGSGSTHGTSLEKHKILHRLLODSSSP 700
Db 3278 ISO---RETESAVSALGSEFGTSSTJDYSLGDGEMSVNELETPPL---VIAPEDE 3330
Qy 701 VDLAKTLTAENGKIDS-----QESSSTAPGEVITIKQEPV---SPKKKEALLRY 747
Db 3331 AALAKALETAGEPASILIEBPMEPERAEPPDEADEIEBPVEVLDPLELNAKV--Q 3388
Qy 748 LLDKDDTDIGLEITPKLER---LDSKTDPASNFKLAMTEKEKMSFEFGDDQSGELD 804
Db 3389 SLKHEDMDIKAD--TPSEKDQIDTDTETERPDAADSSGSLKIDETVQSSSPKKSIS 3446
Qy 805 NIEETILDLOLSQLEPFPTRPGAPAGSVKQALINDIMOLTAENSPTV--PVGAQKTAL 863
Db 3447 NNSPTPRETANIDIPNV--ESQPKLSNESTPOPSTVITKLPLF--DTPKTVAPAGLPSPV 3501
Qy 864 RLSQSTFNN-----PRGQUGR--LLPQNQLPLD----- 890
Db 3502 KIEPTISKLOQPLVQVQVTLPAHSTGSGISANSVINLDSNVISSCSNTSAASATAS 3561
Qy 891 ---TTLSQPTGAGFPPIRNSPSYSVLPQGMNGQGMINGNLGSSGTMGINSASR 946

```



```

Db      3106 KSPSPILCLIGDDDDALHTPAKPTTSSR--GNDGL-----T 3142
Qy      470 PQOSSPGMPPGQPTKSLSRHRMSPCVACSPRIPPSQFSPAGSLSPVVCSTGNSHY 529
Db      3143 PSREKP-----RLISP-----IPKPTIANSSTLSTOSAEPTV-----SSGTVIS- 3182
Qy      530 TNSSLNALQALSEGHVSLGSSSLASPDLMGNLONSPPVMMNPPLSKMKSIDSKOCF--G 587
Db      3183 -SSALATPTTSTTAAGV-----AAP-----GLDNPST-----SASACKKKESTFPG 3224
Qy      588 LYGEPEEGTTGAESSCHPEQKET-----NDPNLPAPVASERADGOSRLHDSKQTK 640
Db      3225 FPGQDDRISESAYOSI-SAEFNSTSLDNINADPKIPAPSPRATKPLDKLEESRNT 3283
Qy      641 LLQLLTTSKDOMPEPPLASSLSDTKNDSTGSLPGSGTHTGSLKEKHKLHRLLODSSSP 700
Db      3284 ISQ-----EETESAVSALLGESFGSTSTTDYSLDGMEMSSVLEETPTL---VIAEPDEE 3336
Qy      701 VDLAKLTAEATGKDL-----QESSSTAPEGSEVTTKOBPV-----SPKKKENALLRY 747
Db      3337 AALAKAITEFACEPASILEEPEREREAPDDPEAEIESEPVVEVLDPELNKAV--Q 3394
Qy      748 LLDKDDTDIGLPEITPKLER--LDSKTDPASNTKLIMKTEKEEMSEPEGDQPSELD 804
Db      3395 SLKHEDMDIKAD--TPOSERLQIDTDEENPDEADSSGSLKIDETVQSSSPERKSTIS 3452
Qy      805 NLEELILDQNSQLPOLFPDTPRGAPAGVCKOAIINDIMOLTAENSPVT-PVGAOKTAL 863
Db      3453 NNSPTPRETANIDIPNV--ESQPKLSNESTPOPSVYTKPLFL--DTPKTVAGLPPSPV 3507
Qy      864 RISQSTFNN-----PRPGQLGR-LLPNONLPLD-----890
Db      3508 KIEPPTISKLOOPLVOPVOTVLPAPHSITGSGISANSVINLDSLNISSCSNTSASATAS 3567
Qy      891 ---TLOSPTGAGFPPIRANSSPYSVIRPQMGNGMIGMGNLGNSTGMIGNSASR 946
Db      3568 ASASISFCSPTAS-----QNMAMPQASTPKQGPITPQOALIRQSLIMQPTISIPQPTH 3621
Qy      947 PPMGEEMAPOSSAVRYTCAATTSAMNRYVQGM---IRNPASTPMRSPPOGOROTL 1002
Db      3622 FAVPMVLSPOS-----HHPOQPGTYWVGIRAS---PHSPILHSRGVAVQ 3664
Qy      1003 OSQVANNIGSELEMMNGGPOYS-----OOQAPRNO-----TAPWESILPIDQAS--FAS 1050
Db      3665 SKLVQOLSP-----VGRPMWSQPSPOOVQOOTQOOHALITTSQSSNISPLASPTRYLS 3718
Qy      1051 QNRQPFGS-----SPDDLICP-HPAESPDEGALLDQLYLALRNFDGLEIDRALGIRE 1104
Db      3719 SENSPTTSKVNSTYORPNQOVPQOPSPKSAVEVOTTPQMLTIPLOKMTPIQVPHH-----PT 3774
Qy      1105 LVSQSQAVDEPEFSSOD-----SNIML-----EOKAPVFPQOYVSAQMAOG 1146
Db      3775 IISKVVTVQPOQATQSOVASPPLISLPRPHKVHNLMAHONQOOPQOVIATKTAHONQOHQ 3834
Qy      1147 SY-----SPMDPRNHTMGORPSYATLRLMORPRGLRPTGLVONOP--NOLRL--- 1191
Db      3835 QPMHOOIMTOROHHOQOOLHGSOOTSA-----FOHMHOOHQAQOQOHHNNOOHLNQ 3889
Qy      1192 -----OLOHRLQAQONRQPLMAQOISVSNVNTLRLRGVTPQARINMAQORRETL 1243
Db      3890 LHAQOHPPOKHQAQOQEFNQOIQOHQSOQ-----QHQQVQOQAQOQHLSQ 3935
Qy      1244 NQHLROROMHOOQOQOQRL-----MMRG--QGLMTPTSMAVPSG-----MPTMSNPR 1290
Db      3936 QOHQSOQOOLNDQOHOQOQOQLOQIQKLOQHMGPOQOQKSPQGVGHLGSTISIFASQOHNQ 3995
Qy      1291 IP-----QANAQOF-----PFPNTYGIS-----QOPDPGTGATTPQSPPLSPMAHTOS- 1335
Db      3996 LPARGVPOQHPOQLSHSPCKPNTLVSVNQGVOVPAITTRGSHSPNOQOOLPHQOSS 4055
Qy      1336 ---PMHQSOQ---ANPAYQAPSDINGMAQNGNGSMFESQOSPPIHFQOQANTSMYSNNMN 1389

```

```

Db      4056 SGHPHOKQLSPGKANLPLQTLPLNVIONTEPKIIYQOHIVAQONVPPROTGANAIIHYQONG 4115
Qy      1390 INVSANITNGGMSNMOMTGOISMTSY-----TSVTSISGLSSMGPEVYNDPALRGULFPN 1445
Db      4116 KD---STPGHVEPPPMASOKTSESVSVIRPTPTTGLAVISANTVGS-LTLEENTLIRI 4171
Qy      1446 QLPQMD-MIKOE 1456
Db      4172 SQPKDELTIED 4183

RESULT 29
QY9VE6 PRELIMINARY; PRT: 2280 AA.
ID QY9VE6
AC QY9VE6
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CG5580 PROTEIN.
GN SBB OR CG5580.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA Anatolides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt J.G., Nelson C.R., Miklos G.L.G.,
RA Abrell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasly E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borrova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferlita S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Gang N.S., Gelbart W.M., Glasser K.,
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mervinov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Relbert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svitskas R., Tector C., Turner R., Venter G.E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
EMBL: AF003800; AAF5721.1;
DB Flybase: FBgn0010575; sbb.
DR InterPro: IPR003880; Znfshpanant_attach.
DR InterPro: IPR000822; Znf-C2H2.
DR Pfam: PF00096; zf-C2H2; 1.
DR SMART: SM00355; Znf_C2H2; 1.

```



QY	91	AANIDEVQKSDVSSGCGVYDKDALGPMLEALDGFEEVFNVBEGNVVFGSENVTOYLRYN	150
Db	90	OSDASEI--RQDKRPP---FLSNEEFTQMLBAGFFLAIMTDGSIIVSEYVSLLEHL	145
QY	151	QDELNNKSVSYSLIHGDTHEFVKNLLPKPSIYNGGSMSEPPR--RNSHTFNCRL-----	203
Db	146	PSDLYDQSIENFIRPGEHSEYVK--ILSHHLEDSLTPEDLKSKNOLEFCCHMLRGITDP	204
QY	204	-----*KKPLPDESEEGHDN--QEAHQ-KYETMQCEA---VSOPKSIK	239
Db	205	KEPSTYEVYRFIGNFKSLNSVSTPHNGFECTIQRTHRSPVEDRYCVCVAIYRLATPOFIK	264
QY	240	EEGEDLOSLICVAVARVAKERPYLPSESTTTQDLOGKTTSLDTSITMGAAKPKGEDEL	299
Db	265	E-----*CTV-----EEP-----NEEFTSHSLMEWFLFLD---HRAAPLIGLPP	302
QY	300	VARCIOKFAHQEGESESVAKRHHHEVLRQCLATSOYRFSLSDGLTVAQOTSKLRSQ	359
Db	303	EVLGTSQGYDHYHVDLESLACHEN--LMQYKKGSCYTYRLTKGQOMIMYIQTHYITYHQ	361
QY	360	TTNEQOLVYISLHML-----HREONCVNMPDLTGOTMCKPLNPISNSPAAQALC	409
Db	362	WNSRREFYVCHTHVSYAEVRAERRELRRELGYESLPE-----TAADKSDQ	405
QY	410	SGNPQODMTLSNINFRPNRGKEDQMPMGKFGSGGGMNHYSGMQATPPGCSYALKMNS	469
Db	406	SGSDNKRITVS-----*LKEALEFFBDS-----PTPASSRSSSKSS	441
QY	470	-PSQSSPQMGNGO--PTMSLSPRHMSPGVAGSPRIPTPOCFSPAGSLHSPGVGCSGTGSH	527
Db	442	HTVAVDPSTPTKLPITDSTRPR-----*PLPRAHE-----KMTQRRS	478
QY	528	SYTNSLNALDALSEGHVSLGSSLASBDLKMGLQNSPYNNANPPUS--KMGSLDS--KDC	585
Db	479	SFSSQSINUS-----QSVSGLTQTPAMSOAANLPTPOGMSQFOFOLSAOLGAMOHKLD-	528
QY	586	FGIAGEPSEGTGQAESCHPGECKETNPDLPAVASEBAGD--QSHLDSKCGQTKL--L	642
Db	529	-----QLEQRTNRIEANIH--KQDEBLR--KIQDQLOMNVHGQGLQMFLOOSNPGCLMGVS	579
QY	643	QLLTTKSD--QMEPSPLASSLSDTNKSDGSLPGSGSTGCTSLKEKHKLHRLLODSSSP	700
Db	580	QLSSGNSNIQOLTPINMGQVVPVNOJOSGVNAGHVS-----GQHMQQQQLOSTSTG	633
QY	701	VDLKLTAELAAGCKLDSOESS--*APSEYTIKQEPVSPKKENALLRYLLDKDPTKDIG	758
Db	634	SOQSMGSHSQPTSLPMQTPBTTLAPLYNTWVISQPAAGSM-----*VP	676
QY	759	LPEITPKLERIDSKTYDPAASNKTLIAKMTKEKEMSFEFGDQPGSELNDLE-----ELLD	812
Db	677	IPSSMPQ-----NSTQSATVYTF-----TQDRQIRFSQGOOLYKLVLYAPACGAVVPST	727
QY	813	LQNSQLPOLPFP--*DRRGAPAGSVDKOAIINDLMQLTAAENSEVTPVAGAKTALRISOST	869
Db	728	MLMGQVYVATPFTATQOQADALSVTQOQ--QOQOQOQOQOQOQOQOQOQOQOQ--QSQQSS	783
QY	870	FNNRPG-----QUGRL--FNQMLPDLITIQSPFGAGPPP	904
Db	784	QDQPHPSVQDPAQLTPPPQFLQTSRLLHGNPSTQILLISAAPLQOSTPP	834
RESULT	31		
ID	Q01YB0	PRELIMINARY;	PRT; 865 AA.
AC	Q01YB0;		
DT	01-DEC-2001 (TREMBLrel. 19, Created)		
DT	01-DEC-2001 (TREMBLrel. 19, Last sequence update)		
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)		
DE	CLOCK PROTEIN.		
GN	CLOCK		
OS	Spalak gallii.		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		

Query Match	4.3%	Score 328	DB 11	Length 865
Best Local Similarity	20.0%	Pred. No. 1,1e-11		
Matches 220	Conservative 160	Mismatches 370	Indels 350	Gaps 53
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Spalacinae;			
OC	Spalax.			
OX	NCBI_TaxID=164323;			
RN	(1)			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=BRAIN;			
RA	Aviary A.;			
RT	"Biological Clock in Total Darkness:The Clock/MOP3 circadian system of the blind subterranean mole rat."			
RL	Proc. Natl. Acad. Sci. U.S.A. 0:0-0(0).			
DR	EMBL: AJ318057; CAC85403.1; -			
SQ	SEQUENCE 865 AA; 97437 MW; E5003191B7578C21 CRC64;			
QY	31 KRNETKRNDEQNKYIEELAEILIFANFNIDNFNRPDKCALIKETVKQIRKDEKAA 90			
DB	39 RAKSEKRRDQFNVLIKELGSMLPGNAR-----KMDKSTVLQKSIDFLR--KHKETIA 89			
QY	91 AANIDEVQKSDVSSTGQGVYIDKDALCPMLLELDGFFPVNLEGNVNVSENVOTQYLRN 150			
DB	90 QGDASEI--RQDKPT--FLSNEFTQLMEALDGFILIMDGSITIVSESTLTLEH 145			
QY	151 QEELMKSVSYSLIHGDHTEFVKNLPRKSTVNGSGWGPFR--RNSHPNCSML----- 203			
DB	146 PSDVDVQSVFNIPGESEHYK--ILSTHLLSDSLTPPEYLSKKNLEPCCIMLKGTYDP 204			
QY	204 -----VKRLPDESEEGHDN--QEAHQ--KYETMOCFA-----VSQPKSIK 239			
DB	205 KEPTSEYVYRFIGNFKSLNSVPTSAHNGEFTQTRHRSYEDRCVAFATVRLAPRFIK 264			
QY	240 EEGEDLQSLICVARRVPRKERVLPSSSEFTRODLOCKITSLDTSTYRAAMKGGMEDL 299			
DB	265 E-----MGTV-----EEP-----DEEFSRSLSEKKFLELD--HRAPPIIGYLPF 302			
QY	300 VARCIOKFFHAQEGESVSAYAKRHNEHVLKROGLAFQIYRFSLSDDGTVAQPKSLINSQ 359			
DB	303 EYLGTSGDYHYAVDDLENLAKCHEN-LMVGKGKSCYRFLTKGOQWIMLYTHYYITTHQ 361			
QY	360 TTNEPOLVLSLML-----HREQNVCMANPDLTGTGKGLNLPISNSPAQALC 409			
DB	362 WMSREFIYCTHTVSYAEVRAERREKJIEBSLD-----AAADKSQD 405			
QY	410 SGNPQODMTLSSNINPFLINGKEQGMGRGGSGGNHNVSGMATTPQGSNTYALKNS 469			
DB	406 SCSDRIRITVS-----LKEALERDHS-----PTPSASSRSRSKRS 441			
QY	470 -PSQSSPGMPCQ--PTSMLSPRHRMSPVAGSPRI--PSQFSPAGSLHSPVGCSSGTGSH 527			
DB	442 HTAVSDPSTPKPIPTDTST-----PPRQHLPA--HEKMAORR----- 478			
QY	528 SYTNSLSNALQALSEGHCVSGLSSLASPDLMKGNLQNSPVNNMPPLS-----KMSGLD 581			
DB	479 SFSQSGMNS-----QSVGPDLTOP--VMSQANLPV---PGMGQFOFQSNQGLAMQ 524			
QY	562 S-KDCEGLYGESECTGOAESSCHPGQOKETNDPNLPRAVSSERRADG--QSRHLDSK-- 636			
DB	525 HLKD-----QLERTHMEIANIH--RQOEELR--KIQEQLQWVHGQGLQWFLQGSNGL 574			
QY	637 --GQKLLQLLTKKDDQEPSPPLASSLSLDYTKNDSTGSLPGSGSTH--GISLKEKKIYLR 693			
DB	575 NFGVQLSSGNSNITQTLTPIMQGVVPTQNOIGSMNAG---HIGTS---QHLIQOS 627			
QY	694 LODSSPYDLAKLTFEATGKGLDQSSSTAGSEVTTIQEVPVSPKKKKMALLRYLLDKDD 753			
DB	628 LQSTS-----TQSQSQSVMSGHSGQOTSLASQO----- 655			
QY	754 TKDIGLPEITPKLERLDKTDPAANTKLIAMKTEKMSFEPGDPQSGELDNLEIIDL 813			
DB	656 -----STLTAPLYNTVYIS-----QPA--PGSMV----- 677			

Qy	814	QNSOLPOLFEPTRRGARAGSVDMQAIINLMOLTAEINSVYTPVGQAKTALRTISOSTFNNP	873
Db	678	---QIPSSMP-----QNSTQSAITYTF--TQDRIRSQ-----	706
Qy	874	RPQGLGRLLPNQNLPLDITLQSPYTGAGPEPPINSSPYVTPQPGMGNQMGINQNLG	933
Db	707	---GOOLVYK-----LVTAPACGAV-----MVESTLMGCV-----	735
Qy	934	NSTGCMGNKSNRPTMPSGEMAPQSSAVRYTCAATTSAMNRPVQGMININPAASITPMRPS	993
Db	736	-----VTATPTEATPQQAQATLSV-----TQQAPOQ-----QQPQQAQPPQQAQ	773
Qy	994	SQPGQRQTLQSQVNNI--GPESELENNMGPPQYSQQA-----PPNQTAWPESILPDIQAS	1047
Db	774	PQQAQQAQSSQEQQLRQSVYPPQSAQLQTSPPQQLFTSLNLGNSTQLI--LSAAPFLQGST	831
Qy	1048	FASQNRQDPFGSSPDDLCPH	1067
Db	832	FPFHHQQAQHQSSQQAQQLSRH	851

ID	Q9UGC3	PRELIMINARY;	PRT;	5533 AA.
AC	Q9UGC3;			
DT	01-MAY-2000 (TREMBLrel. 13, Created)			
DT	01-MAY-2000 (TREMBLrel. 13, last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, last annotation update)			
DE	SPEN RNP MOTIF PROTEIN LONG ISOCORM.			
GN	SPEN OR CG18497.			
OS	Drosophila melanogaster (Fruit fly).			
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;			
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
OC	Ephydroidea; Drosophilidae; Drosophila.			
OX	NCBI-TaxID=7227;			
RA	[1]			
RC	SEQUENCE FROM N.A.			
RP	STRAIN-IS01;			
RA	McLiette E.L., Harding K.W., Mace K.A., Ronshaugen M.R., Wang F.Y.,			
RA	McLiette W.;			
RT	"Spenn encodes an RNP motif protein that interacts with Hox pathways to			
RT	repress the development of head sclerites in the Drosophila trunk.";			
RL	Development 0:0-0(2000).			
DR	EMBL; AF188205; AAF13218.1; -.			
DR	HSSP; P09651; 1HA1.			
DR	FLYBase; FBgn0016977; spen.			
DR	InterPro; IPR00504; RRM.			
DR	Pfam; PF00076; rrm; 3.			
DR	SMART; SM00360; RRM; 3.			
DR	PROSITE; PS50102; RRM; 3.			
SQ	SEQUENCE 5533 AA; 597114 MW; AFC606DE06DDEF269 CRC64;			
Query Match	4.3%;	Score 328;	DB 5;	Length 5533;
Best Local Similarity	20.2%;	Pred. No. 1,6e-10;		
Matches 330;	Conservative 197;	Mismatches 655;	Indels 450;	Gaps
OY	8	TSDSRAETRRKRCPCPOLGSPKRNTERKNNRQENKYLELAELIFANFNIDINENPK	67	
DB	2819	TSAPTASTQTSRKG-----	EDKMEFLFGIISDESSQF	P 2851
OY	68	DKCAILKETVYQIQIKEQEKAAANIDEQKSDVSTGGVADKDALPMLLELDGFF	127	
DB	2852	EQA-----ETNK-----	DIIPSSVSTT-----	GPIVSALQ--- 2877
OY	128	FVNVLEGNVVFSENVTOYLRYNOELMKNVSYSILHV-GDHTFVKMLLPRKSIVNGS	185	
DB	2878	-----	TYKEPSTPRSKNEAHIGIQLTVHEPPQQQLRSLRSLSGS	2911
OY	186	WSGEPRRNSHTFNCRLIVPLPDSEEGHDNOEAHQKTYEMOCFAVSPKSI--KEGE	243	
DB	2918	SSSSHADERRERRRKR--EKRRKRSQREGQNO-IHKSQKVT-KVDDNSVDVDEAGR	297	

0Y	244	DLCLCLVARRPMKRPVLPESEFETTRKODLQKITSJLDTSTWRAMAKGWEIYLVR	303
Db	2974	ALEKQL-----MSDFTPKLYSE-----EATSTATKTRSDMT-----	3006
0Y	304	IQFHACHEGESVSYAKRHHNEVLROGLAFSQIYTPRSLSDGTLYAQAOTKSKILRSQTTNE	363
Db	3007	VERFSDNEDNNSV-----DMTKGVXSEQOEJHKSXD-----KKKKKKRKEKQ	3051
0Y	364	POLYISLHMLHREQNVCMYMRDLTGOTMK-----PLNPTSSRAHQALC	409
Db	3052	EKL-----LQOORRESLPNVASTSAPPTPKCLVYVQAASKHADQLDLAKHIXSP---PVC	3105
0Y	410	SGNGQDMTSLSNINFPINPKEQOMPMRGFEGSGGGMHNHVGMAOTTPQGSNYALKMNS	469
Db	3106	KPSSLCLCLDODDDALHTPKAKPTTPSSR--GNDGL-----T	3142
0Y	470	PSQSGPMNPQOPTSMLESPHNRMSPGVAGSPRIPTPSQSPAGLSHPVGCSSTGNSHSY	529
Db	3143	PSREKP-----RLTSP-----IPKPTIANSSTLSTQSAETPV---SSQTVIS-	3182
0Y	530	TNSLSMLQALBSHGIVSLSSGLSLASPDLMKGNLONSPPVMMNPPLSKMGSLDKSCF--G	587
Db	3183	-SSALATTPTTSTRAAGS-----AAP-----GLDMSPT-----SASAQCKKKESTFPG	3224
0Y	588	LYGPSEGGTGOAESCHPGEOKET-----NDPNLPVAVSERADQSRHLDHSGKGT	640
Db	3225	FDGQDLDRISBSAVQSI-SAEFNSTSLINDIADPEKIPVAPSPRATKPLDKLEESKSHVT	3283
0Y	641	LLQLLITKSQDMERPSSLASLSDTKNDSTGSLPSGSSHTGTSLKKKHILHRLDSSP	700
Db	3284	ISQ-----EETESAVALLGSEFTGTSSTDYSLDGMDEMSVLELPTL---VIAEPDEE	3336
0Y	701	VDLAKLTAEATGKDL-----OESSSTAQSEVITKQRPV---SPKKENALURY	747
Db	3337	AALAAKALETAGBAPSLLEEREMERERAPRDPDAELSESPVVEVLDPELKNAY--Q	3394
0Y	748	LLDKDTRKDGLPETTPKLER---LDSKTDPSANTKLIAMKTEKEMSEFREDQGSGLD	804
Db	3395	SLKHEMDMDIKAD--TFQSERQDIDTDEENDEADSGSPILKIDETVOSSSPKXSIS	3452
0Y	805	NLEFIIDLQNSQOLPFPPTRGARVAGSDKQALINDLMQTAENSPVT--PVGAKQKAL	863
Db	3453	NNSPPEPATNIDIPNV--ESQPKLSNESTPQPSVTTKPLP---DTRKYPAIGLPSPV	3507
0Y	864	RISOSTENN-----BRPOLGR-LILPNQNLPLD-----	890
Db	3508	KIEPPTISKLOOPLVCGVQVULPAPRSTGSGISANSVIMDLNSVYSSCNSMTSASATAS	3567
0Y	891	-----ITLQSPGAGPPPIINSSPYEVIQPGMMGQMGIGNOGMLGNSSTGMIGNSAR	946
Db	3568	ASASISFSGSPAS-----QMAPQASTPQGGITPQQAIRQSLIMOPRTTISTEOTPH	3621
0Y	947	PTMSEGEAPQSAVRYTCATTSANMRPVQGM---IRNPAASIPMRSSQPGQROTL	1002
Db	3622	FAYQOMLSPQS-----HHNQRETYWVGIRAS---RPHSLHSPERGVAQ	3664
0Y	1003	QSOYMNIGPSELENNMGPOYS---QOQAPRNO-----TAPRESJLIPDOAS--PAS	1050
Db	3665	SRLVGQOLSP-----YGRPMVQSPSPQOOVOQOQOONALITIPROSINISPLASPTTVLS	3718
0Y	1051	QNRQPFCS---SPDQLLQ-HPAAESPDSDEGALLDOLYLHNNQDGEIEDRALMIRE	1104
Db	3719	SSNSPTTISKVNSIYQPRQOVPQOQSPKSVAEVQTTQOLMTPLQKMPITQVPHN---PT	3774
0Y	1105	LVSQSOAVDEPQFSSOD-----SNIML-----EQKAVFPRQOVLASQOAMQ	1146
Db	3775	IISKVTVYQROAQATQSOVASSPPLGSLRPHKHNVLNANOQOQPOVAKMTAHNOQIMQ	3834
0Y	1147	SY-----SPQDPNFTMQQRSPYATLHMQRPBGILRPGVLQVONR---NQLRL---	1191
Db	3835	QFMHQOMIQRQOHHQOQOOLIGOSQOULTSA-----RQCHMOOHOAQOQOONHNOQOLHQ	3889

```

OY 1192 -----OLHRLQAOQNROPPLMNOJISNVSNVLTLPBGPVQAPINAQMLAORRELT 1243
DB 3890 LHAQOHPTQKHQAQOQOQNOQIQHQSOQ-----QHVOQOQNAQOQHLSQ 3935
OY 1244 NQHLRQRMHQOQOQYQOQTL-----MMRG-QGLMTPSMVAPSG-----MPATMSNR 1290
DB 3936 QOHQSOQOOLNOHQOQOQOQLOQIQKLOQMHGPOOQKSPQGVGHGSGTSTIFASOQHSNQ 3995
OY 1291 IP-----QANAOQF-----PPPPNYGIS-----QOPDPGTCATTPQSPMLSPRMAHTOS 1335
DB 3996 LPARVPOQOHPQOLSHSPCKSPKPTLVSNQGVOPPALITRVGSHSQPQOQOQLPHQOSS 4055
OY 1336 ---PMMQSO---ANPAYQAPSDINGMAQNGMGSMPQSOQSPPHGQOANTSMYSNNMN 1389
DB 4056 SGHPHQKQLSSPGANLPLPTPLANTVQNTPKTIYQOHIVAQNVPPRQGNALHTRPONG 4115
OY 1390 INVSATVTGKSSMNQMTGQISMTSY-----TSVSTGLSSGPPQVNDPALRGNLPPN 1445
DB 4116 KD---STPPGHVEPTPPAMSAQKTSSESVIRTPPTTGLAVISANTVGS-LITTEENLKI 4171
OY 1446 QLPQMD-MIKOE 1456
DB 4172 SOPKODELIED 4183

```

```

RESULT 33
OY1YB2 PRELIMINARY; PRT; 865 AA.
AC OY1YB2;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CLOCK PROTEIN.
CN CLOCK.
OS Spalax carmeli.
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Spalacinae;
OC Spalax.
OX NCBI_TaxID=164324;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA Aviyi A.;
RT "Biological Clock in Total Darkness:The Clock/MOP3 circadian system of
RT the blind subterranean mole rat.";
RL Proc. Natl. Acad. Sci. U.S.A. 0:0-0(0).
DR EMBL; AJ318058; CAC85404.1;-.
SQ SEQUENCE 865 AA; 97419 MW; CA58A56F535C6625 CRC64;

```

Query Match 4.38; Score 325; DB 11; Length 865;  
 Best Local Similarity 19.98; Pred. No. 1.7e-11;  
 Matches 219; Conservative 161; Mismatches 370; Indels 350; Gaps 53;

```

OY 31 KRNTKRRNEQENKYEELIIFANFDIDFNFKPKCALIKETVQKIROIKQOEKAA 90
DB 39 RNKSEKRRDDQFNVLKRLGSLPBNAR-----KMKSTYLOKSIDPLR-KIKETITA 89
OY 91 AANIDEVOKSVSTGQGVIDKDALGPMLEALDGFVVNLNENGVNVSENVTOYLRYN 150
DB 90 QSDASEI-RODMKPT---FLSNEEFTQMLEALDGFPLAIMDGSIIYVSESVLSLEHL 145
OY 151 QEELMKNKSVSILVGHDTVEVKNLPRKSIYNGSGWSEPPR-RNSHFNCML----- 203
DB 146 PSDLVDSVFNFIPEGHSEVYK-ILSTHLESDSLTPRYLKSQKQLECCMHLGCTIDP 204
OY 204 -----VKPLPDESEEGHDN--QEAHO-KYETMQCFA-----VSQPKSIX 239
DB 205 KEPSTYEVVRFLGNFKSLNSVPTSAHNGFEGEIQTRHPSYEDRCVATVYLAIPQITK 264
OY 240 EBGEDLOSLCIVARRVPMKEBPVLPSSSEFTTRDLOGKITSLDTSTWRAAMKQWEDJ 299
DB 265 E-----MCTV-----EPP-----NEEFTSRHSLMKFLFLD---HRAPIIGYLPF 302

```

```

OY 300 VRCIOKFFHAQEGESVSAYAKRRHHHEVLROGLAFSOIYRFSLSDGTLVAQTKSLIRSO 359
DB 303 EVLGTSGDYTYHVDDELNLAKCHEH-LMVGKQKCYTRFLTKGQOMIMLOTHYITYHQ 361
OY 360 TTNEPOLYISLML-----HREQNVCMVNDPLTGQTMGKPLNPSSNSPAHQALC 409
DB 362 WNSRPEFIVCTHTVSYAEVRAERRELIGIESLDP-----AAADKSQD 405
OY 410 SCNPQODMTLSSNINFPINGKEQMGMPRGKGGSGGMNHSGMQATTPQGSNTYALKANS 469
DB 406 SCSDNRINTVS-----LKEALERDHS-----PPPSASSRSRKS 441
OY 470 -PSOSGPMNPGO-PTSMLSPRHRMSPVAGSPRIIPPSQFAGSLHSPVGCSTGNSH 527
DB 442 HFAVSDPSTPTKIPITDST-----PPROHLPA-----HEMAQORS----- 478
OY 528 SYTNSLNLALQALSBGHVGLSSLASPDCLKNGNLQNSPVNNPPPLS-----KMGSLD 581
DB 479 SFSQGSNMS-----QSVGPSLTQPIYISQA--ANLPV---PQMGSGQFQSAQLGAMQ 524
OY 582 S-KDQFGLTGEPSSEGTGQAESCHPEQKQETNDNPLPPAVSERRADG-QSRHLDSK--- 636
DB 525 HLKD-----OLEQRTMIEANIH-ROQEELR--KIOEQLOMVHGQGLQMFLOQSNPGL 574
OY 637 --GQTKLQLLTKSDQMEPSPLASSLSDPTNKDSTGSLPGSGSTH-GTSLKKEKHLIHL 693
DB 575 NFGSVOLSGSNSTIQQLPIMMGQVPTNQIQSGMAG---HIGTS---QHLIQOOS 627
OY 694 IODSSSPVDLAKLTAEATGKDLQOESSSTAPGSEVTIKQEPVSPKKENALLRYLIDKD 753
DB 628 IQSTS-----TQSSQGSVMGSHSQSTSLASQO----- 655
OY 754 TKDIGLPEITPKLERLDSKTDPASTKLIAMKTEKEMSFEFGDPGSELONLEETLIDL 813
DB 656 -----STLTAPLVNTWYIS-----QPA--PGSV----- 677
OY 814 QNSQLPQLEPOTRRPAPAGSVQKAIINDMLQJTAENSPVTPVGAQKATALRSQSTFPNP 873
DB 678 ---QIPSSMP-----ONSTQSAVTTF--TQDROIRESQ----- 706
OY 874 RQQLGRLLPNQNLPLDITLQSPGTAGPFPPIRNSSPYSVIRPOPMGNGMIGNGLG 933
DB 707 -----GQQLVYK-----LVTPAPVACAV-----MVSITLMQV----- 735
OY 934 NSTGMIGNSASRPTMPGEGWAPQSSAVRYTCATTSAMNRPVQGMIRNPAASIPMRPS 993
DB 736 -----VTAYPTFATQOQQAQTLV-----TQOQPOQ---QOPOQOQPOQOQ 773
OY 994 SQPGOROTLQSGVNNI-GPSELEMMMGCPQYQOQA-----PPNGTAPWRESILRIDAS 1047
DB 774 PQOQOQSSQOEQQLPSVPPQSAQLQSPQOFLQTSRLHLHGNSTOLI--LSAFLQOST 831
OY 1048 FASQNRQPFSSPDDLLCPH 1067
DB 832 PPSHHQOQSOQOQOQLSRH 851

```

```

RESULT 34
OY1YB8 PRELIMINARY; PRT; 865 AA.
AC OY1YB8;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CLOCK PROTEIN.
CN CLOCK.
OS Spalax judaei.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Spalacinae;
OC Spalax.
OX NCBI_TaxID=134510;
RN [1]

```





```

OY 233 SPSKKEGEDLQSLICVARRVPMKERPVLPSSSEFTTRDLOKITSLDTSTRAM 292
DB 249 ATPQIKR-----MCT-----VEESNEFTSRSLKMKFLFD---HRAPP 286
OY 293 KRWEDLVRCIOKPHAOHEGESVYAKRHHNHEVLROGLAFSOIYRFSISDGTVAQAOK 352
DB 287 IIGYLPFEVLGTSYGYVHVDLEMLAKCHEH-LMOYGKSKCYRYFLTKGGOWIMLQTR 345
OY 333 SKLISQTTNEPOLYISLHML-----HREONVCYMANDLTGO----- 389
DB 346 YITTHQNSRBEFTVCTHTVSYAEVGAERREKRNEDSPPAITAEKNQDSVDNHNMT 405
OY 390 -----TMGKPLN-----PISNSPAHQALC----- 409
DB 406 VSLKALERFDSDSRTPSSSKSSIKSSHTAVSDPSSTPTKIPDTISTPRQALTLGLDKR 465
OY 410 -SGNPGQDMLTSSNINFPINGP-----KEOMGMPGRFGS-GGMNHVSG----- 452
DB 466 RSSISSQSMS-SQSVSQPLSQSVMKOTASIQLOQGTQMPQFTAQFGAMKHLKQLEQR 524
OY 433 -----MQATTPOGSNTALKNMSPSSPGMN--PGOPTSKMLSPR 489
DB 525 TRIEENIORQOELRKIODQLOMVGOGIOMYLQ-----QPARGLNFGPVQYSSGNSPS 579
OY 490 HRMSPGVAGSPRIIPSPSPAGSL---HSPVGVCS--TGNSH-----SYTNSLNAALQA 539
DB 560 IQQLP-----QFTMGQGVVOTNOLQGVNMTHGVGAQHIMQOQQLHNTSQOQGN 628
OY 540 LSEGHG--VSLGS---SLASPLMKGNLONSPVNMNPPPLSKMGLSDKDCGLGEPS 593
DB 629 IHGHNQGTSLSSQTSGLTSP-----LYNTMVISQRP----- 661
OY 554 EGTTOAESSCHPGQKETNDPNLPPAVSSERADGOSRLHDSKGQTKLIQLT 647
DB 662 SGSVMQSMNIQOQSNQ-----GASVITFPQDRQIRF--SOAQOIVTKLVTT 705

RESULT 36
OY 091906 PRELIMINARY: PRT: 825 AA.
AC 091906:
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE CLOCK.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=RETINA;
RX MEDLINE=20153706; PubMed=10686352;
RA Zhu H., Larue S., Whiteley A., Steeves T.D., Takahashi J.S.,
RA Green C.B.;
RT "The Xenopus clock gene is constitutively expressed in retinal
RT photoreceptors."
RL Brain Res. Mol. Brain Res. 75:303-308(2000).
DR EMBL: AF227985; AAF34772.1;
DR InterPro: IPR001092; HLM_dmc.
DR InterPro: IPR003015; HLM_dmc.
DR InterPro: IPR001067; Nucleinslocator.
DR InterPro: IPR001610; PAC.
DR Pfam: PF00989; PAC; 1.
DR Pfam: PF00989; PAC; 2.
DR PRINTS: PR00785; NCTRNSLOCATR.
DR SMART: SM00353; HLH; 1.
DR SMART: SM00086; PAC; 1.
DR SMART: SM00091; PAS; 2.
DR PROSITE: PS00038; HELIX_LOOP_HELIX; UNKNOWN_1.

```

```

SQ SEQUENCE 825 AA; 93269 MW; 0F657C3D9C5F24E0 CRC64;
Query Match 4.2%; Score 321; DB 13; Length 825;
Best Local Similarity 21.0%; Pred. No. 2.8e-11;
Matches 172; Conservative 118; Mismatches 262; Indels 266; Gaps 34;
OY 31 KRTEKRNREOEKYEELAEELIFANFNIDNENFRKPCALIKETVVKOIRQKEOEKA 90
DB 30 RAKSEKRRDQNLILKEIGSMIPGNARI-----DKSTVLOKSIDFLQ--KHKEISA 80
OY 91 AANIDEVQKSDVSTGQVYIDKDALGPMLEALDGFEEVYNLEGVNVESENTOYLRYN 150
DB 81 QSDASEI-RQDKMPT---FLSNEFTQMLLEADGFEFLVMTGNTIYSESVTSLEHL 136
OY 151 QEELANKSVYSLIAGDHTPEYKNNLPKSIYVNGSGMSGEP--RNSHTNCRNL----- 203
DB 137 PSDVDQSIFFNVPDEGSEHYK-IISTRMLESGLSISEYLAETKRNLEFCCHLRTQADP 195
OY 204 -----VRPLPDSEEGHD---NOEAHQKYETMOGFA-----VSOPKSIK 239
DB 196 KEPSTIEYKFTIGNEKSLNNVFPSTNGFDGALQSLRPREYERVCVATVRLATQFIK 255
OY 240 EGEDLQSLICVARRVPMKERPVLPSSSEFTTRDLOKITSLDTSTRAAKPMGMDL 299
DB 256 E-----MCT-----VEESNEFTSRSLKMKFLFD---HRAPP1IGYLPF 293
OY 300 VRRCIOKPHAOHEGESVYAKRHHNHEVLROGLAFSOIYRFSISDGTVAQAOKSKILRQ 359
DB 294 EYLGTSYGYVHVDLEMLAKCHEH-LMOYGKSKCYRYFLTKGGOWIMLQTRYYITYHQ 352
OY 360 TTNEPOLYISLHML-----HREONVCYMANDLTGO----- 389
DB 353 WNSRPFIVCTHTVSYAEVGAERREKRNEDSPPAITAEKNQDSVDNHNMTVSLKEAL 412
OY 390 -----TMGKPLN-----PISNSPAHQALC-----SGNPGQ 415
DB 413 ERFDDSRTPSPSSSKSSIKSSHTAVSDPSSTPTKIPDTISTPRQALTLGLDKRRSSISQ 472
OY 416 DMTLSSNINFPINGP-----KEOMGMPGRFGS-GGMNHVSG----- 452
DB 473 SMS-SQSVSQPLSQSVMKOTASIQLOQGTQMPQFTAQFGAMKHLKQLEQRTRIEEN 531
OY 453 -----MQATTPOGSNTALKNMSPSSPGMN--PGOPTSKMLSPRHRMSPGV 496
DB 532 IORQOELKRIQDQLOMVGOGIOMYLQ-----QPARGLNFGPVQYSSGNSPISQQLP-- 584
OY 497 AGSPRIIPSOFSFAGSL---HSPVGVCS--TGNSH-----SYTNSLNAALQALSEGHG- 545
DB 585 -----QFTMGQGVVOTNOLQGVNMTHGVGAQHIMQOQQLHNTSQOQGNIHGCHNQ 635
OY 546 -VSLGS---SLASPLMKGNLONSPVNMNPPPLSKMGLSDKDCGLGEPSDEGTTGQA 600
DB 636 QTSLSQTSQGLTSP-----LYNTMVISQRP-----SGSMVQM 668
OY 601 ESSCHPGQEKETNDPNLPPAVSSERADGOSRLHDSKGQTKLIQLT----- 647
DB 669 PSNTIQOQSNQ-----GASVITFPQDRQIRF--SOAQOIVTKLVTTPMACGTVMVPSIM 718
OY 648 -----KSDMEPSPLASSLDTNKDSTGSLP 673
DB 719 FMGPVVTAVPTFTTQQOQFQTLSTFHQOQNOQOQVYP 756

RESULT 37
OY 09W7C3 PRELIMINARY: PRT: 875 AA.
AC 09W7C3:
DT 01-NOV-1999 (TREMblrel. 12, Created)
DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE CLOCK.

```

OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OC Gallus.  
OX NCBI\_TaxID=9031;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=RETINA;  
RX MEDLINE=99337784; PubMed=10407173;  
RA Larkin P., Baehr W., Semple-Bowland S.L.;  
RT "Circadian regulation of iodopsin and clock is altered in the retinal  
degeneration chicken retina.";  
RL Brain Res. Mol. Brain Res. 70:253-263(1999).  
DR EMBL; AF132531; AAD3283.1; .  
DR HSSP; P36956; 1AM9.  
DR InterPro; IPR001092; HLH dim.  
DR InterPro; IPR003015; HLH\_Myc.  
DR InterPro; IPR001067; NucleusLocator.  
DR InterPro; IPR001610; PAC.  
DR InterPro; IPR000014; PAS.  
DR Pfam; PF00785; PAC; 1.  
DR Pfam; PF00989; PAS; 2.  
DR PRINTS; PR00785; NCTRNSLOCATR.  
DR SMART; SM00353; HLH; 1.  
DR SMART; SM00086; PAC; 1.  
DR SMART; SM00091; PAS; 2.  
DR PROSITE; PS00036; HELIX\_LOOP\_HELIX; UNKNOWN\_1.  
SO SEQUENCE 875 AA; 98725 MW; 04DDEBBD79747A4 CRC64;

Query Match	4.28;	Score 319;	DB 13;	Length 875;
Best Local Similarity	19.28;	Pred. No. 4.1e-11;		
Matches 189;	Conservative 137;	Mismatches 363;	Indels 294;	Gaps 37

```

0y 31 KRFNFKRRENNKYIEELALIELIANFNDIDNFNFKDKALIKYQJROIKEQKAA 90
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 39 RNSKSKRRDDPVLVILKLGSMLEPGRN-----KDKSVLQKSIDFLR--KHKEITA 89

0y 91 AANIDEVOKSDVSTGGVIDKALGPMMLTALDGEFFVYNLEGVNVFSENYOYLRYN 150
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 90 QSDASEI--RQMKPL---FPSNEEFTQJLMEILADGFLAIWTDGNIIYVESVYPLEHL 145

0y 151 QEELMNKSVYILHVGDTHEFVVKLLRSLIYNGSGMSGEPFR--RNSITFNCRL----- 203
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 146 PSDLDQSVFNFIEGEHSEIYK--ILSSHLESDLSLPEYLKSNQNEFCOHMIRGTIDP 204

0y 204 -----VKPLSEEECHDN--OEANO-KRETMQCF-----AVSOPRSK 239
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 205 KEOPTEYEVKTIQDNFKCLNNVPNSAHNGEFTIORSHPREDEKVCYIATVRLATPFIK 264

0y 240 EGEDLOGLCIVARRPMKPERVLPSESEFTTRDLOGLTISLDTSTYMRANKPMEDL 299
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 265 E-----MCTV-----DEP-----NEEFTSRHSLEMKPLFID--HRAPITGLPF 302

0y 300 VARCIOKHAQHEGESVYAKRHHHEVLROGLASQIYRPSLSGTLVAQOTKSLRSQ 359
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 303 EYLGTSQGYDHYHVDLDLNLACHEH--LMQYKGRKSCYRFLTKGOOMIWLQTHYIYHQ 361

0y 360 TTNEFOLVILSHML-----HREONVCMVRDITLQGMKPLMBISNSPAHQALC 409
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 362 WNSRREFLYCHTYVYTAEVAERRELRRELIGIESLPEI-----398

0y 410 SGNPQCDMTLSSNINFPINGRKEQMGMPGRFGGSGGKNHVSQMOA-----TTPQGS 461
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 399 KADKQSD-----SGSDNHINTVSLKEALEREFTSPTS 431

0y 462 NTKLMKNSPQSSSPCMNNGOPTSMLSPRHMSFGVAGSPRIPRQSFAGLSHPVGYCS 521
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 432 ASSRSRSKSHTAVSDHDSSTPKM-----TYDSTPRPO-----SLSAHE-----K 472

0y 522 STGNHSTYNSJNLQALQSLSGHGVSLSSSLASPDLMKGNLQNSPVMNPPPL-----SKMG 578
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 473 STORSSSLSSOSLSS-----QSLQRPVOTQPMOPATLQHOSSKQBPVQFQSAQLG 523

```

[illegible]

RESULT	38
Q9W6Q2	
ID	Q9W6Q2
PRELIMINARY;	
PRT;	853 AA.

```

DE DE
OS Gallus gallus (chicken).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RM SEQUENCE FROM N.A.
RX MEDLINE=20490740; Pubmed=10931848;
RA Chong N.W., Bernard M., Klein D.C.;
RT "Characterization of the chicken serotonin N-acetyltransferase gene.
RL Activation via clock gene heterodimer/E box interaction."
DR EMBL, AF144425; AD032860.1; -.
DR HSSP; P36956; IAW9.
DR InterPro: IPR001092; HLH_dim.
DR InterPro: IPR003015; HLH_MYC.
DR InterPro: IPR001610; NucleusLocator.
DR InterPro: IPR000014; PAS.
DR Pfam; PF00785; PAC; 1.
DR Pfam; PF00989; PAS; 2.
DR PRINTS; PR00785; NCTRNSLOCATR.
DR SMART; SM00353; HLH; 1.
DR SMART; SM00086; PAC; 1.
DR SMART; SM00091; PAS; 2.
DR PROSITE; PS00036; HELIX_LOOP_HELIX; UNKNOWN_1.
SQ SEQUENCE 853 AA; 96297 MW; 5349C5C1F7293C97 CRC64;
```

Query Match 4.2%; Score 318; DB 13; Length 853;  
 Best Local Similarity 21.7%; Pred. No. 4.5e-11;  
 Matches 165; Conservative 116; Mismatches 249; Indels 232; Gaps 37.

D	b	39	RAKSEKKRRDQPNVLIKELSGMLBGNR-----KMDKSTVLQKSIDFLR--KHKXITA	89
Q	y	91	AANIDEVOKSDVSSGTQGVIDKDALGRPMLEALDGEFFVNVLEGNVVFSEMTQYLRYN	150
D	b	90	QSDASEI--RODMKPT---FLSNEEFTQIMLEALDGFALMTIDGNITIVSESVPLLEHL	145
Q	y	151	QOELMKNKSVSYILHVGDTHEFVNKILRKSYLNGSGNSGEPFR--RNSITFNCRL-----	203
D	b	146	PSDLVDOSVFNFITQGEHSELYK--ILSSHLLESLSLTPELYLKSNNQOEFCHMLRGITDP	204
Q	y	204	-----VKRPLDEEERCHDN--OEANO-KYETMQCF-----AVSOPKSIK	239
D	b	205	KEQPILEYVKFIGNFKCLANNPNSAHNGFBETIDRSHRPSTEDKVCYIAIYRLATPOFIK	264
Q	y	240	EEGEDLOCLICVARARVPMKERVLPSSSEFTTIPRODLOGKITSLDTSTMRANKPGMEDL	299
D	b	265	E-----MCTV-----EEP-----NEEFTSRHSLEMKFLFD---HRAPPIGLPF	302
Q	y	300	VRRCLQKHAOHEGESYAKARRHNEVLROGLAFSOLYRPSLSGTLVAAQOTSKLIRSQ	359
D	b	303	EVLGTSGDYHNVDDLNLAKCHEN--LMOYKKGKSCYCYRELTQOQOIMLOTNYHYIYHQ	361
Q	y	360	TTNEPOLYISLHML-----HRONQCVNMPDL-----TGQIMGR--LNPJ-----	398
D	b	362	WNSRPERFVCTHTVYSTAENRAERRELDGLEESLPETIKADKSDQSGSDHNIHTVSLKEAL	421
Q	y	399	-----SSNSPAHOAL-----CSGNPGQDMTL-----	419
D	b	422	ERFDTSPPTASRSKRSKSSHTAVSDHSSPTKMTVDTSTPRQSLSAHHEKSTORSSLS	481
Q	y	420	-----SSNINEPPIKPE-----QMGMPGRFEGGS--GGMNHY-----	450
D	b	482	SQSLSSOSLQPVQPTMSOPATQLQSSMSQFSAQLGAMQHLKDLEORTMIEANIR	541
Q	y	451	-----SGMATTPQ-----GSNY--ALMKNPSPOSS-----PGANPOQPTS	484
D	b	542	QOEBLRTIQEODLIVHOGQLOMFLQOSTSGLNFSSVDTLGTGSSSSYQOALRPGMMQGOVVO	601
Q	y	485	MLSPRHMSPCVAGSPRIIPRPSQFSPACSLHSPPVGVCSSTGNSHSYTNSSLNALQALSEGH	544
D	b	602	TNORQSGMNGHTSTPRHIDQO-----PL-----QSASQNHQON-----VLSGH	611
Q	y	545	GVSLGSSLASFDLKMGLQNSPVVMNBPPLSKMGSLSDKDCFGLYGEPSGTTGQAESSC	604
D	b	642	GOQ--SSLA-----GOSQNT--VSTPLNTM-----VISQPTAGNVQVYVSSSL	680
Q	y	605	HREGQETNPENLPRAYSSEADQSLNHSKQGTIKLLOLT	646
D	b	681	-----PONNNN--AAAVTTFQDQOIF--SQGOULVTKTKLT	714

RESULT	39			
Q9X7A4				
ID	Q9X7A4	PRELIMINARY;	PTG;	870 AA.
AC	Q9X7A4;			
DT	01-NOV-1999	(TREMBLrel. 12, Created)		
DT	01-NOV-1999	(TREMBLrel. 12, Last sequence update)		
DT	01-DEC-2001	(TREMBLrel. 19, Last annotation update)		
DE	ENDOTHELIAL PAS DOMAIN PROTEIN 1/HYPOXIA-INDUCIBLE FACTOR-2 ALPHA.			
GN	EPAS1/HIF2 ALPHA.			
OS	Bos taurus (Bovine).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;			
OC	Bovidae; Bovinae; Bos.			
OX	NCBI_TaxID=9913;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=ARTERY;			
RX	MEDLINE=99255430; PubMed=10320777;			
RA	Hara S., Kobayashi C., Imura N.;			
RT	"Molecular cloning of cDNAs encoding hypoxia-inducible factor (HIF)-1alpha and -2alpha of bovine arterial endothelial cells.";			

```

RL      Biochim. Biophys. Acta 1445:237-243(1999).
DR      EMBL, AB018399; BAA78676.1; -.
DR      InterPro: IPR001092; HLH_dim.
DR      InterPro: IPR003015; HLH_MYC.
DR      InterPro: IPR001067; Nucleinslocator.
DR      InterPro: IPR001610; PAC.
DR      InterPro: IPR000014; PAS.
DR      Pfam: PF00785; PAC; 1.
DR      Pfam: PF00989; PAS; 2.
DR      PRINTS: PR00785; NCTRNSLOCATR.
DR      SMART: SM00353; HLH; 1.
DR      SMART: SM00086; PAC; 1.
DR      SMART: SM00091; PAS; 2.
DR      PROSITE: PS00038; HELIX_LOOP_HELIX; UNKNOWN_1.
SQ      SEQUENCE   870 AA: 96168 MW;  FEE602EE6012D7712 CRC64;

```

Query Match	4.18;	Score 314.5;	DB 6;	Length 870;
Best Local Similarity	19.58;	Pred. No. 7.6e-11;		
Matches 215;	Conservative 147;	Mismatches 348;	Indels 395;	Gaps 44

OY	10	DPSRAETRRKRCSPROLGSPKRNTEKRRRROENKIYEELA-ELIFANRNDIDNFKRD	68
Db	6	EKKRSSEERRKE-----KSRDARCRSRSETEFEVYLAHELPLPH-----SVSSHLD	52
OY	69	KCAILKETVQKQJROIKEOEKAAANIDEVOKSDVSTGGVIDKD-ALCPMILEADJGF	127
Db	53	KASIMRLAISFLRTRK-----LSSVSCENSEADADQOMNLVYKALEGT	99
OY	128	FVNVLGNAVVEENVTOYLARYNOBELMKNKSVYSILHVDHTEFVKNLLPKSVNGSMS	187
Db	100	AVVOTDGDWIFLSENMISKRMGLTQVELTGHSLFDFTPCDHEIRENL--SILKNSGRG	156
OY	168	GEPPRRNSTHFCNRYLRLPDSEEBGHDNOZAHOKIETMOCFAVSPKSIKEEG-----	242
Db	157	-----KSKSMSTERDFEMKCKCTVTNNGRTVNLKSAWKV	192
OY	243	-----EDLOSLICVAVRRVPKREVRPLR-SESFTTRDLOG	278
Db	193	LHCITQVQVYNNCPRHSSLSICKEPLSLILLMCEPIQMRSHMDITLDSKTEPLSRISM	252
OY	279	KITSLDSTMRAPMKPGW--EDLVRCRLOCFHQAHEGESVYAKRRHNEVLRQGLAFSOT	336
Db	253	KFTYDGD---RITELVGYHPEELLGRSAEFYHALDSENMNT--KSHQMLCYKRGQVVSQ	306
OY	337	YRFSLSDGTVLAAQKRSKILIRSQITNEPOLVLSLH--MLHRQONCVAMPDLTGOTMGPR	394
Db	307	YRMLAKHGGYVWLETOGVITYNPRMLQPCIMCVNYVLSLEIKENDVVFEMDOT-ESLFRP	365
OY	395	-----LNPISNSPAAHQALCSGNPDGMTSSNINPFGPKOMGMPGRFGSGGGMNHY	450
Db	366	HLLTNNSLIFDNS-----GKVAVSEKSNLEFTYKLBEE-PELLQOLAPTAGDTII	412
OY	451	SGMOATTP--OGSNYALKMNSPSSQSPGMNPGOPTSMLSPKRRHMSPGVAGSPRIIPSO	507
Db	413	S-LDGETNFNEBESSAYGKGLIPRPGQWGWGEVAKSHGT-----H	448
OY	508	SPAGSLHS-PVGVCSSTGNSHSYTSSLNALDALSEBGHVSLSGSSLASDLMKGNLQNSP	566
Db	449	SEAGSLPFEVYPOAALGNSTPSASS-----SSCTSPSSPGDYVTSID	492
OY	567	VMMNPPPLSKMSLSKDOFGLYGEPSEGTGOAESCHPGOKETND--DNLPRPAYS	623
Db	493	DNLKTEALEKLFAMDT-----EAKDQC--GTOTDNEJDLTTLAPYIM	534
OY	624	BRADQO-----SRLHDSKGQTKLLQLLITTKSDQMEP--SPASSLSIDNKDSTGLSPGS	675
Db	535	DOEDPOLSPICEESLBPETPQSAPOHCHPSTMSNIFQPLAPMAS-----	578
OY	676	GSTHETSLKEKHKILLRLLDSSSEVYDLAKLTAEATGKDLQSOESSSTAP-----	724
Db	579	---HSTFLFDKKO-----QOLESKKTEBEQRRVSTAFED	609

```

QY 725 -GSEVITIKQ-----EPVS-----PKKENALLRYLLDKDT 754
DB 610 GGSRSVSLQCCGQTYRPLSMGSGISNTQWPPDPDLQGLPTKMGEDRHA-----SAVGA 653
QY 755 KDIQLEIRPKLERLDSKT-----DPA-----SNTKLIAMKTEKEMSFEP--- 795
DB 664 APLGLPPLPPLHMLMKRKSCKGFGPGCPDYMSAMATLSKLLKQLQLEEDAFQDMG 723
QY 796 GQPGSELNLE-ELIDDLQNSQLPOLFPDTRPGAPAGVDAKQAIINDLMQTAENSPVT 854
DB 724 GQPGSGTSHLMWKRMKSLRGSGTSLMPDK----- 754
QY 855 PVGAQKALRISOSTFNNRPPGOLGRLLENQNLPLDITLQSPGAGPPPIRNSPSYVI 914
DB 755 -----LPNANVPNDEFIONPV-RGRSOPPLRHLSP----- 782
QY 915 POPGMNGMGIMGNQNLGSSGTMGINSASRPT---MPSGEWAPOSSANRYVCATTS 971
DB 783 POP-----PSAT-----SPEPTKSGFPQCYAPOTDYSLPAHMKSG 821
QY 972 MNRPVGGMIRNPASIRMPRSSOP 996
DB 822 M-----ASRLLGSPFEP 833

RESULT 40
QVEG7
ID 09VEG7 PRELIMINARY; PRT: 2703 AA.
AC 09VEG7
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE OSA PROTEIN.
GN OSA OR CG7467.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidae; Drosophilidae; Drosophila.
OX NCBI_Taxid=7227;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RA MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Aamathides P.G., Scherer S.E., Li P.W., Hoskins R., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortan J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.L., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Flosser C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostali D., Houston K.A., Howland T.J., Wei M.-H., Ibeagham C.,
RA Jafali M., Kalush F., Kapen G.H., Ke Z., Kemtison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasse P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mervulov G., Mishina N.V., Modary C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Murzy D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleby J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Rehner K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

```

```

RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Massaman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou W., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL: AE003718; AAF55457.1; -.
DR Flybase: FBgn0003013; osa.
DR Interpro: IPR001606; ARID.
DR Interpro: IPR002965; P_rich_extensn.
DR Pfam: PF01388; ARID; 1.
DR PRINTS: PR01217; PRICHTEXTSN.
DR SMART: SM00501; BRIGHT; 1.
SQ SEQUENCE 2703 AA; 282894 MW; 27ED5BA0277A6369 CRC64;

Query Match 4.1%; Score 314.5; DB 5; Length 2703;
Best Local Similarity 20.1%; Pred. No. 3,9e-10;
Matches 245; Conservative 111; Mismatches 475; Indels 387; Gaps 48;

QY 392 GRPLNPIS-----SNSPAHQALCSGNPGQDMTLSSNINFPIN--GPKQGMGMPMRFG 443
DB 676 GPPRPPTSOAGACGANSMPGSAQAGTGGKMPNHTGQIPRYQWVPSPQOTVPGAGPG 735
QY 444 SCGMNHYSGMQA-----TPPGSNYALKNMSPSQ-----SSP----- 475
DB 736 AMGNHYGCGKTPRPVYGGPRPPGSGSPRLNYLKHQHGKGGVGSPTPPGPGY 795
QY 476 ----GNNPGOPTSMLSBRHMSPGVAGSPRIIPSGFASGL--HSPVGCSTGNSHS 528
DB 796 NGPTGMNPGMP--MGPRNHMGPRHGTNNMPPTSTPPOSQMLQGGCPQCGASG-- 848
QY 529 YTNSSLNALQALSEGHVSLG--SLASPLDKMGNLQNSVYNNNPPRLMSGLSDKDF 586
DB 849 ---PESGGRHISQDNQISSGPTGAGAHNAVTSVTTGPDGTSMDVSSQSTLSNNSAA 905
QY 587 GLYPERSEGTGQAES-----CHPG-----BQKETNDNPLPAYS 622
DB 906 S--GDPQCTPRSKRNNDPYSQSHLAPRSTSPRVYMHGCGGEEDMSSPNWPRAG 963
QY 623 SERADGOSRLHDSKQGTLLQLLT--KSDQMERPLASSLSTPNKDSITGLSGSGSTHG 680
DB 964 SPQVFN---HVPVQPEPFKSTITTKKSDSL-----CKLYEMDNDPRRGMW----DKLR 1011
QY 681 TSKKEHKILHRLDQSSPVDAKL-----TAAETGKDS-----QESSST 722
DB 1012 AFMEERKPTTACPTTSKQPLDLYRLIYKKEGGEVEYTKSTWKDIAGLIGASSSA 1071
QY 723 APGSEVITIKQEPVSPKKENALLRYLLDKDTK-----DIGLEIRPKLERLDS-- 771
DB 1072 A-----YTLRKHYTKNLLTFECHFDRGIDIDPLRIIQVEAGS 1108
QY 772 --KTPASNTKLIAMKTEKEMSFEPEDDQSGSELNLEELIDDLQNSQLPOLFPDTRPGA 829
DB 1109 KKTTAAASVSPSGSSNSQDSFPAPPGSAPMAIDG-----YGPYGGGS 1152
QY 830 P-----AGSVKQALINDLMQTAENSPVYVGAOKTALRISOSTFNNRPPGQGLPLPNQ 885
DB 1153 PLYVASPPQPDYTAGOMQRPSPONNPOT-----PHFGAAVAACD 1194
QY 886 NLPLDITLQSPGAGPPPIRNSPSYVIPQPG--MMQMGIMGNQNLGNSGTMGNSA 944
DB 1195 NISVSNFEDPIPAAGGPGSGTGRPGQGGPRGAASGAAVAVAG-----GCPQ 1244
QY 945 SRPTMBSGEWAPOSSAVRYTCAATTSAMNRPVYOGMIRNPASIRMPSSQPGQROTLAG 1004
DB 1245 PHPPRP--SPHT-----AAQQAAGQHQQHNPQHNGHGLGPRPPRQOQCGG--QQ 1292
QY 1005 QVANNIGSELEMMGGRQYISQQAAPRQGTAPWES--ILPIDQASFASQNNQPRGSSPDDL 1063
DB 1293 PPSVVG-----GGPPAPQOHGPGQVPPSPQOHVRAAGAPY-----PGGSGYPTP 1339

```



**THIS PAGE BLANK (USPTO)**



Query Match	Best Local Similarity	100.0%;	Score 7631;	DB 19;	Length 1464;
Matches 1464;	Conservative	0;	Mismatches	0;	Gaps
1 MSGMGENTS	SDPSRAETRRKRECPDOLGSPKRNTERKNEQNKYIEELAEIFANFNDI	60			
1	msmgentsdp	raetrkrkecpdqlgspktrntekrneknyieelaelfanfnndi	60		
61	DNFENRPRKCALIKETVYKQIRQIKEDQKAAANIDEVQKSDVSVTGGQYIDKDALGPMML	120			
61	dnfnkprkcaliketvkvqirgikedeqaakaanidevqkdsdvstggvldkdalgpmmll	120			
121	EALDGFEEFVNVLGGNVVFESENVTOYLARNQELMNKSSVYSLIHVGDHDEFPYKNLLPKSI	180			
121	ealdgfeefvvnlggnvvfeseenvtoylarnqelmnkssvylshvgdhdefpynllpkssi	180			
181	VNGSGWSGEPFRPNRSTHTFCRMIVKPLPDSSEEGHNDCAHQKRYEMQCEFAVSOPKSI	240			

Db	181	vngsgvsepprrmshfncrmlvkprpdseeeghndgeahqkylmcgfaxsqpksike	240
Qy	241	EGEDJQSLICVARHVPKMERPVLPSSSEFTTRBQDLQKITSLSDTSTMRAAMPQMEDVJ	300
Db	241	egedjgsclicvarhvpkmerpvlpsssefttrqdlqgkltstdtmaamkpyweli	300
Qy	301	RRCLOKFLAHCHEGSSVYAKRHHHEVLRGLASQLYRRSLSIDGTLVAAQTSKILRBOT	360
Db	301	rrclqkflhaqhegessyakrhhhevlrqglatsiqlyrrtsldgtlvaqtkskilrtqt	360
Qy	361	TNEPOLVYSLMLREONCVCMNPDLTGCMKELMPISNSPAHQALCSGPGDMLIS	420
Db	361	tnepolvyslmlreoncvcmpndlgtcmkelpnisnsparhualcsnpggdmlls	420
Qy	421	SNINEPIMGKMGMGKFRGSGGGMNHVSGQATPDGSMVALKMSPSQSSFGMNP	480
Db	421	sninelpimgkmgmgkfrsggggmnhvsgqatpdgsmvalkmspsqssfgmnp	480
Qy	481	QPTSLMSPRHRMSPGVAESPRIPEPQSPAGSLHSPVGCSSSTGNSHSTYNSLNAOL	540
Db	481	qptslmsprhrmspvgvaspripepqspagslshspvgcssstgnshtytnslnaol	540
Qy	541	SEGHVSVLSSSLAPDLKMGNLONSYPVNMNPPRLSKMSLSDKOCRGLYGESEETQOA	600
Db	541	seghvsvlssslapdlkmgnlonsypnmnprrlskmslstdkocrglygepseetqga	600
Qy	601	ESSCHPGOKRTNPNLPVAVSSPRADGGSRLHDSGOKRLLOLTTKSDQMEPPLASS	660
Db	601	esschpgokrtnpnlpvavsspradggslhdsygokrlloltkdsdqmepplass	660
Qy	661	LSDTNKDSTGSLPGSGSTHGTSLKEKHKLHRLLODSSPYDLAKRLAETATGKLSQSS	720
Db	661	lsdtnkdstgslpgsgsthtslkekhlhrrlldsspydlakrlaetatgklsqss	720
Qy	721	STARCEYTTIQOEVSFKKEMALRXLDKDTRKIGRPEITPKLERLSDTDPASNTK	780
Db	721	starpceyttiqlqoevskkrenalrlylxlkddtkdgiqlpeitpklerlstdtspasntk	780
Qy	781	LIAKTKEEKSFPBGPOPSSELDNLEELIDLONSOLPOLDPRPGAPASVYKAOIT	840
Db	781	liakteekesfepbgpopseldnleelidlonsolpoldprpgapasvykagii	840
Qy	841	NDLMQTLAENSPLYVGAOKTALRISQSTFNNRPGOLGRLLPNOLPLDITLOSPTAG	900
Db	841	ndlmqltaensplyvgaoktaltalrisqstfnnrpgolgrllpnolplditlosptag	900
Qy	901	PPPIRNSSPYSVTPOGAMGNOGITGNOGNCNLSSTGIMGNASARPTMPSGEMAPOSA	960
Db	901	pppirnsspysvtpogamgnogitgnogncnlsstgimgnasarptmpsgemapossa	960
Qy	961	VRYVCAATTSAMNRPVGGIMRNPASITPRPSSQGOGOTLOSOMNIGSELEMMNG	1020
Db	961	vrvvcaattsamnrvpggimrnpaasitprpssqgogotlosomnigselemmng	1020
Qy	1021	POYSOOQAPMPOATPWESEILTPIDQASFASONKOPRGSSPDDLLCPHRAESPDEGALL	1080
Db	1021	pyysooqapmpoatpweaseltpidqasfasonkoprsgsspddllcphraespsdegall	1080
Qy	1081	DOLYALKNEDGEEIDRALGIPELVQSQOAVDEQFSSQSDSNIMLEOKAPVFOQVASQ	1140
Db	1081	dqlyalknedgeeidralgipelvqsqoavdeqfssqsdsnimleqapvfpqvyasq	1140
Qy	1141	AQMAGSGSPMODNPFMTMKGORPSYATLRKQRPGRGLPRFGLYONQPNOLRLQLOHRLQAO	1200
Db	1141	aqmagsgspmodnpfmtmkgorpsyatlrkqrpgrglprfglyonqpnolrlqlqhrllaq	1200
Qy	1201	ONROPLKMOISNVSNVNLTLTRPGVPOAPRLNAOMLQRORELINQHLRQRMHMOOQOYOQ	1260
Db	1201	onroplkmoisnvsnvnltrpgvpoaprlnaomlqrorelinqhlrqrmhmqoqqoyq	1260
Qy	1261	RTLMKRGGLMTPESMVAFSGEMATMSNPRIPDANAQOEPFPNPTGISQOQDPGFTGATY	1320



Db 1261 rllmmrggglmlmtspmvapgmptlmsnprilpqaagqfipfpayisqppdftgatt 1320  
Oy 1321 POSPLASPRMAHTOSPMMQSOANPAYOAPSDINGMAOGNMGNSMPQSPPHGGOAN 1380  
Db 1321 ppsplmsprmatlqspmmqgsanpayaqpsdlnqwaqmgngsmfsgqspbhqgan 1380  
Oy 1381 TSMYSNNNNINVSMAATNTGMSNNQMTGQISMTSVTSVTSGLSSMGPEQVNDPALRG 1440  
Db 1381 tsmysnnnninvsmaatntgmsnnmqmtgqismtsvtsvtsglssmgpeqvndpalrg 1440  
Oy 1441 NLFPMQLGCMIMIKOGSTTRKVC 1464  
Db 1441 nlfpmqlgcmimikogsttrkvc 1464

## RESULT 2

AAM99482  
ID AAM99482 standard; Protein: 1463 AA.

AAW99482;

08-JUN-1999 (first entry)

Murine NCoA-2 protein.

Mouse; P/CIP; p300/CBP/co-integrator-associated protein; gene expression;  
breast cancer; inflammatory disease; atherosclerosis; osteoporosis.

Mus sp.

Key location/Qualifiers

Misc-difference 989 /label= unknown

XX MO9856806-A1.

XX 17-DEC-1998.

XX 12-JUN-1998; 98WO-US12263.

XX 12-JUN-1997; 97US-0049452.

XX (REGC ) UNIV CALIFORNIA.

XX Glass CK, Rose DW, Rosenfield MG, Torchia J;

XX WPI: 1999-080883/07.

XX New nucleic acids encoding P/CIP and NCoA-2 polypeptides - are used  
XX to identify agents that regulate gene expression, e.g. for treatment  
XX of cancer, inflammatory disease and osteoporosis

XX Claim 22; Fig 2A; 100pp; English.

XX This sequence represents the amino acid sequence of the mouse NCoA-2.  
XX The protein can regulate gene expression so are potentially useful  
XX therapeutically, e.g. against (breast) cancer, inflammatory disease  
XX (e.g. atherosclerosis) or osteoporosis.

XX Sequence 1463 AA;

Query Match 78.2%; Score 5969.5; DB 20; Length 1463;

Best Local Similarity 80.5%; Pred. No. 0; Mismatches 163; Indels 33; Gaps 9;

Matches 1191; Conserved 95; Mismatches 163; Indels 33; Gaps 9;

Oy 1 MSGMGENTSDPSRAETRRKECPDOLGSPKRNTEKRNREDENNKYEELAEITAFANFDI 60  
Db 1 msgmgentsdpsraetrrkecpdolgspkrntekrnredeknyeeelaefanfdi 60  
Oy 61 DNFNKPKCAILKETVQAIQIKOEKAAANIDEVOKSDVSTSGGVIDKDALGPMUL 120  
Db 61 dnfnkpkcailketvqaiqikekkaaanidevqkdsstsggvidkdalgpml 120

Oy 121 EALDGEFFVNLNENWVSENVTOYLRYNOBELANKSVYSLIHYGDHTEPVKNLLPKSI 180  
Db 121 kalldgeffvnlnevwseenvtoylrynobelanksvysslhygdhtepvknllpkxsm 180  
Oy 181 VNGGSMSEPPRRNSHTNCRMLVPLPDSEEGHDNDQAKYETMOCFAVSQPSIKE 240  
Db 181 vnggsmsepprrnshtncrmvplpdseeghdndqakyetmofavsqpsike 240  
Oy 241 EGEDLOSLIT-----CVARVPKEREVLPDSSEFTTRDLOGKITSLDNST 287  
Db 241 egedlqslitvmedphgktsliirk-lynpqpprdmftghymme-----sr 290  
Oy 288 MRAAM-KPGWEDLVRCLOKFEHQEGESVSRAKHHHEVLRQGLAFQIYRFSISDCTL 346  
Db 291 heagigresgkkafr-----stshim-ksalympirrhnevrlrglatrsqyrlfslsdqcl 346  
Oy 347 VAAQTKSKLIRSQTNBEPOLVLSLHMLHREQNVCVANPDLTGOTGKPLNPISSNSPAHQ 406  
Db 347 vaagtksklirsqtneqvlvislhlhreqmncvmpdltgqngkplnpiassspahq 406  
Oy 407 ALCGSPGQDMTLSSNINFPINGKEQMGMPGKFGSGGGMNHWGMOATTPQGSNVALK 466  
Db 407 alcsqnpqdmhlsishlfpingkeqghpgrfgsggmngvsgahqattpqgsnyalk 466  
Oy 467 MNSPQSSPGMNPQGTMLSPRHRMSPVAGSPRIPPSOPSPAGLSHPVGVCSGTNS 526  
Db 467 mnspsqsspgmnpqgavslsprgrmspvagsprlppsqsipaglnmspvvcsstgns 526  
Oy 527 HSYTNSLNALQALSEGHVSLGSSIASPDLKMGNIQMSPVMMNPPSLSKMSLSDSCDF 586  
Db 527 hsythslnalqalseghvslgssiaspdlkmgnlqmspvmmnppslshgslsdscdf 586  
Oy 587 GLYGESECTTGQANSSCHPGQKETTNDPMLPRAVSSERADQSRHLDKSGOTKLQLLT 646  
Db 587 glygesektgqakaschpkkqkphdsnpqasgraeasrlhdsqgkllqllt 646  
Oy 647 TRSDQMEPSPLASSLSDINKDSTGSLPGSGTHGSLTEKKHILHRLDSDSSPVDLAKL 706  
Db 647 tksdqmepsplsslsdnhkdstgslpgsgthgsltkkhlhrlldsdsspvdlakl 706  
Oy 707 TAATGKDLQSESSSTAPGSEVTIKQEPVSPKKNALRLVLDKDDTKDGLPEITPKL 766  
Db 707 tahatgkdlqseessstapgstvtikqepvspkknallrlyldkddtkdglpeitpkl 766  
Oy 767 ERLDSKTPASNTKLIANKTEKESFEPGQPGSELNLEBIILDDLONSQLPQLFPTTR 826  
Db 767 erldsktpasntklianktivkeevsfepdqpsgsklnlikliddlqhsqqlfpttr 826  
Oy 827 PGAPAGSVDKQATINDLQTLAENSPPVPVGAOKTALRISQSTFNNPRLPGQLRPLQN 886  
Db 827 pgapagsvdkqatindlqtladssppvpaqqaakalchssstfnnpprqqlrplqn 886  
Oy 887 LPLDITLQSPGAGFPPIRNSSPYSYVPOPGMNOGMINOGNIGNLSSTGMIGNSASR 946  
Db 887 lplditlqspgagpfpirnspsysvlpqgghngmimgnlsqgnlghsctghysgstr 946  
Oy 947 PTMPSGEWAPQSSAVRVCAFTSAMNRPVOGMR--NPAASIPMRSSPGGORILOS 1004  
Db 947 ptmpsgewapqssavrvcaftsamnrpvogmr--npaasipmrsspggoriqlqs 1004  
Oy 1005 QVMNIGPSELEMMNGGPYSSOQOAPPNOTAPWPSTIIPIDQASRQORPFGSSPDLL 1064  
Db 1006 qvmnigpselehmnggpyssqoapppnotapwpsitlipidqasrtqorpfpgsspdll 1064  
Oy 1065 CPHPAESPDEGALLDQLYALRNFDLEIIDLALGPELVSSQAVADPQFSSQSDNI 1124  
Db 1065 cphpaespsdegalldqlyalrnfdeidldalgpelvsqavadaefssqsksi 1124  
Oy 1125 MLEOKAIVFPQOYVASQAOAGSYSPMODPNFHTMGORPSTATLRMOPRGLRPTGLVON 1184  
Db 1126 hleqkavfpqyasqaghaaggyhpbqdpnfhtmgprpnyctlrhprpqlrptglvgn 1184





```

Db      357  tndrhgfvschfifgrengnyrpnbpvgqgllrpfmagcnssvvg-----mmsmpqglqnp 412
Qy      421  SNINPFINPKEOMGMPMGFFGSGGMNHVS---GMOA--TTPOGSNVYALKMNPSSQSPG 476
Db      413  sstrayladrpstcgmsgaraygssniaaltgpgmgpspsysgmnyglmmspphspg 472
Qy      477  MNGOPTSMLSPRHMSPGVAGSPRIIPSOFPDAGSLHSPVGCSTGNSHSTYNSLAN 536
Db      473  lapnqgnlmisprnr-----gsplksnhqfispvagvhsmaesgntgn- hsfssslsa 525
Qy      537  LQALSGHGVSLGSSLASPDLMKGNLQNSPVNMPPLPSMGSLDSDKDFGLXGE--PSE 594
Db      526  lqaisegvtsltsltsppgk---ldnsp-mnmlcpksvsnqdskspljfydcdpnpve 561
Qy      595  GTTGOAESCHPEGEKJETDNPMLPRAVSSERADGOSRLHDSKQOTLLOLTLTKSD---- 650
Db      582  ssmcgnsrdrhisd-keskes-----svegaenqrgplskghkhlldltsddrgh 634
Qy      651  -OMEPSPLASL-----SDTNKSTGSLPGSGSTHGTSLKREKHLRLLDSSS 699
Db      635  ssitnspldssckessvsvtlpsgvasstsgvssnmhgsllgkchrllhkllqngns 694
Qy      700  PYDLATLTAATKDLSDSSSTAPSGEVTIKEPVSPPKKE-NALLRYLLDKDTRKDIG 758
Db      695  paevakltaeatgk---tsstscgdgnvkvqgqlspkkennallrylldrddpsdal 751
Qy      759  LPEITPKLERLSDKTDPASNTKLIAMKTEKE-EMSEFPGDQPSGLDNEEILDDQNSO 817
Db      752  skelqgvqevdkmqsccsstlpsssgkdkiktetseesgddndaldgdltsed 811
Qy      818  LPQLPDRTRGAPAGSVDKOAILINDMOLTAENSPVTPVGAQKTALRISOSTFNBRPGO 877
Db      812  fymslssn-gshlgt--kqvvtg-----gtneislksesqv--qslrtp-- 850
Qy      878  LGLHLNQNPLD--ITIOSPTGAGPPPIRNSPYSVIP-QGMMGNQGMIGNQNLGN 934
Db      851  -----pytravslspsvsgsppvknsaifmrlpkgmllqnpnmtdsgenys 900
Qy      935  SSTGMIGNSASRPTMPSGEWA-POSSAVRYTCAATTS-----AMNRPOGGMIRN 983
Db      901  smggprnrtvtctpsgdglpnskaqgmepmnsmtgpgdyntslsprlalsgi-- 958
Qy      984  PAASIDMPRSSOPGQRQT--QSOVNNIGPSELEMMNGGPOYSQOQAPNQTAPWESIL 1041
Db      959  -plbtrslslpgarvylqggqgmrlmqrpgeilmqmganpyr-qaaasnqlgswpdgm 1015
Qy      1042  PIDQAFASQNRQPRFSSPDLLCPHPAESPSDEGALLDQLALARNPD--GLEEDRA 1099
Db      1016  smeqvshgicnprllrnslddlvgppsnlegsgderallldqhlntlsntaalgledra 1075
Qy      1100  LGIPELVSOQAVDPEQ--FSSQDSNIMLEOKAPVPPQOYASQOQAQGSYSPMODNFI 1157
Db      1076  lglpelvngqgallepqdafiggeaavmmdqkaglygqfypagppmgq-----gfh 1127
Qy      1158  TMGQRPSTVATLRMO-----PREGLRPTGLVO---NOPNQLRLQLHRLAQO---QN 1202
Db      1128  lqggspsfsmnmqmgqgnfplqgmhpranlmrptcnlpkqlrmqlqglqggqfing 1187
Qy      1203  RQPLMOISN-VSNVNLTLRPGV-PTQAPFINQMLAORQELLNOHLR----- 1248
Db      1188  rgalclkmepntaggaavmrpmmqpgqgflnaqmvagrstellsbhftrqrvammngqg 1247
Qy      1249  -----QROMQOQOVQORTLLMNRGQGLMTRPSNVAPSGMPATMSNRPQANOQFPR 1301
Db      1248  qggqggqggqggqggqggqggqcgafesppnvtaspsmdglllqgplmpagppqdfy 1307
Qy      1302  PPNYGISQOPDPGATGATTPQSPFLMSPRMAHTQSPMWOQSQANPAYQAPSDINGMAQGN 1361
Db      1308  qpnygmngqppdaefgyrvsppnammasrmngpsqnmqmhpbqaaaiygs--semkywpsgl 1366
Qy      1362  GGNMSFSQGSPPHFGQOANTSMTSNNMNTINVSMAITNTGKSSNMOMTGOISMTSVTSVST 1421

```

```

Db      1367  arnsfsfq---qfahqgnpavys-----mvhnmgsghngmnmnmpm 1408
Qy      1422  SGLSSMGPEQ 1431
Db      1409  sgm-pmgpdq 1417

RESULT      5
AAB59278
ID  AAB59278 standard; protein; 1412 AA.
XX
AC  AAB59278;
XX
D7  27-MAR-2001 (first entry)
XX
DE  Human SRC-3 protein.
XX
KW  Steroid receptor coactivator-3; SRC-3; antisense; infection;
XX  Inflammation; tumour; cancer.
XX
OS  Homo sapiens.
XX
PN  US6156571-A.
XX
PD  05-DEC-2000.
XX
PF  15-NOV-1999; 99US-0440612.
XX
PR  15-NOV-1999; 99US-0440612.
XX
PA  (ISIS-) ISIS PHARM INC.
XX
PI  Bennett CF, Cowser LT;
XX
DR  WPI; 2001-079549/09.
XX
PT  Novel antisense compound useful to prevent or delay infection,
XX  inflammation or tumor formation, specifically hybridizes with and
XX  inhibits the expression of human steroid receptor coactivator-3 -
XX
PS  Example 13; Column 43-54; 36pp; English.
XX
CC  The present invention relates to an antisense oligonucleotide,
XX  targeted to a nucleic acid molecule encoding human steroid receptor
XX  coactivator-3 (SRC-3). The invention is useful for inhibiting the
XX  expression of SRC-3 in human cells or tissues in vitro. It is
XX  useful for diagnostics, therapeutics, prophylaxis and as
XX  research reagents and kits. It is useful prophylactically, to
XX  prevent or delay infection, inflammation or tumor formation.
XX
SQ  Sequence 1412 AA;

Query Match      35.5%; Score 2711; DB 22; Length 1412;
Best Local Similarity 42.3%; Pred. No. 2.9e-168;
Matches 642; Conservative 242; Mismatches 439; Indels 194; Gaps 46;

Qy      1  MSGMGENTDPSRAETRRKKECPDQLGPPKRTKERNRNEQEKYTEELAEILFANFNDI 60
Db      1  msllgenl-dplasdskrkklpodtrpgglltcsgekrirregeeskyieelaelsanlst 59
Qy      61  DNEFNRPDKCALIKETVKOIRQIKQEOKAAANIDEVQKSDVSTGQYIDKDALGPMML 120
Db      60  dnfnvprdkcaalketvtrqgkqgk-tisndddvqkavssstgvyldkslpglll 118
Qy      121  EALDGEFFVYNLEGNVFEVSENVTOYLRKNQOEELAMKSVSYLHVGDHREPKNLPKSI 180
Db      119  qaldfglfivndgnaivfsvenvtgylykqgedlvntsvynllheedrkdflkn-lpkst 177
Qy      181  VNGGSGSGEPRRNSHTFRCRMVLRPLDSEEGHNOEAOHKYEMQCFAYSQPSKISE 240
Db      178  vngvswtnetqrgkshfncrmkmktrphlledlnaspeimrgiryetcmqcfalsqprame 237

```





PT steroid receptor coactivator-1, particularly useful for preventing,  
 PT delaying or treating infection, inflammation or tumor formation -  
 PS Claim 1: Column 45-54; 36pp: English.

CC The present invention relates to an antisense compound of up to 30  
 CC nucleobases in length, which specifically hybridises with and inhibits  
 CC the expression of human steroid receptor coactivator-1 (SRC-1) (also  
 CC known as F-SRC-1 and NCOA-1) gene. The antisense compounds are useful  
 CC for diagnostics, therapeutics, prophylaxis, or as research reagents or  
 CC kits. The antisense oligonucleotides are useful for treating a disease,  
 CC particularly a human, suspected of having or being prone to a disease  
 CC or condition associated with the expression of SRC-1. In particular,  
 CC the antisense oligonucleotides are useful for preventing, delaying or  
 CC treating infection, inflammation or tumour formation. The present  
 CC sequence is human SRC-1.

XX Sequence 1441 AA;

Query Match 28.3%; Score 2159; DB 22; Length 1441;  
 Best Local Similarity 36.0%; Pred. No. 4.2e-132;  
 Matches 579; Conservative 235; Mismatches 426; Indels 368; Gaps 61;

QY 1 MSGMGENTSDPSRAETRRKRECP-DOLGSPKRNTEKRNREKNYIEELAFANFMD 59  
 DB 1 msgjgdsdpanpsdhkrkgspscdtlas-----tekrregeenyleelaelsaaisd 56  
 QY 60 IDNRFKDKCALIKETVKQIRKOEKAAANIDENOKSVSTGGCVIDKDALGPM 119  
 DB 57 IdsIsvkpdkcklkktdvqldqImkrmeqskstdddvqskdsissqgvtelkeslgrll 116  
 QY 120 LEADGFEFVVNLEGNVSEPNVQYLRNOEELMNSVSLHVGHTFEVVKLTPKS 179  
 DB 117 lealdgffivncogrlyfveenvtlygynqealmntsvsllhvqdaetvknllps 176  
 QY 180 IVNGSGMSGEPRRNSHTFNCRMLVLPDSEEGHNOEAHQKYEKOCFAVSOPKSTK 239  
 DB 177 lvngvwpqeatrnshtfncrmllhp-pd-epgteneaqeqryemqcfvsgpklsq 233  
 QY 240 EEGDLOGLCVARVYMKERPVLPSESEFTTRDLOGKITSLDTSTIRAAKMGWEDL 299  
 DB 234 edgedfgqclclartlrp--tppaltgyeslmtkqdtctgkllsldtslraagrtgswedl 291  
 QY 300 VRRICQKHAQHEGESVYAKRHHHEVLROGLAFQIRFSLSDGTVAAGTKSLRSQ 359  
 DB 292 vrkclayaff-qpgyregysarqqlfgevmtlrgtaspsyrflindgtcmshntckclcyppq 350  
 QY 360 TTNEPQVLYSLHMLHREONVCMNP-DLTGOTMGR-----LNPISNSPAHQALCSGN- 412  
 DB 351 spdmqpfimglhldrehs--glspqddtngsmisprvnpvnp--slspahgyarstsl 406  
 QY 413 -PGQDMLTSSINFPINPKEO-----KGMPMGRGGSGGMNHVSGM-----QA 455  
 DB 407 ppsnsmnstrlin-----rqsdsdlhssshnsnsnsgsfqspqslvanvalnkqqa 460  
 QY 456 TTPGGSNVALKMSNPQSGSPGMNPGQPTSMLSPRHRMSPGVAGSPRIPSOFP-AGSLH 514  
 DB 461 ss-qskspshlnmpmegtgislaq---fmsprtrvtslgtlatrprimnnsfppristcls 516  
 QY 515 SPVGVCCS--TGNSHSTYNSLNLALQALSEGHSVLSGSSLASPDLMKGNLSPVNMNP 572  
 DB 517 spvgmtssacnnmrnsynlptslqgmegnpnsvfsgsaasprlrmnsgspertlrl 576  
 QY 573 PLSMGSLDSKDCGVLGEPSEGTTCGAESSCHPECKET-----NDPMLPRAVSSERADG 628  
 DB 577 p-aaeskdndkeiasclnemi-----gsdnssdqkpldsqllhndel1-----sdg 622  
 QY 629 QSRLHDSKGQKTLQLLTTTKSD-OMEPSPILASSLD-----TKKDSGSLPGSGSTHG 680  
 DB 623 dskY--qstshklyqlllttaeqglrhaditckdvtscgttsaanssgsgcspsh 680  
 QY 681 TSLKEKHKILHRLIQDSSSPVDLAKLTAEATGKDLSSSSSTAPG--SEVTIKQEPVSP 737

DB 681 seltarhklhlrlige-gpsdlttltsvpeddkd-satsvsvtqvgvnskilelds 738  
 QY 738 KRKE--NALRLYLDKDD-----TKDIGLPETPKLERIDSKTDPASTKLIAMKTEKE 789  
 DB 739 krkeskdhqlllylldkdekdlrslnlslddvkvvekke-qmdbcenlpmckrpe 797  
 QY 790 EMSFERGDDPGSELDLEETLDDLON-SQDPQLFPDTRRGARAGSVDKQAIIINDLMQLTA 848  
 DB 798 eikleaqsqftadldqfdqlppltlekaaqipglcetdrnd--gavtsvlixsel----- 849  
 QY 849 EMSVTPVGAQKTALRISGSTFNPNRPQGLRLPNQNLPLDITLQSPGAGFPPIRMS 908  
 DB 850 -----lpslqsat----- 858  
 QY 909 SPYSYIPOPMMGNOGMIGNOGNLGNSTGMIGNASRPTMPSGEWAPOSASAVRYCAAT 968  
 DB 859 -----arp----- 861  
 QY 969 TSAMNRPVGGMIRNPASIRPSPSQPOGROTLDSQVNVNIGPSELEMNGPQISQQA 1028  
 DB 862 tsrlnr-----lp-----elelaingfygpyt 885  
 QY 1029 PENQAPMP-ESILPIDQAFSAQNRPQSGSPDDLCPHRAESPSDEGALLDOL--YL 1085  
 DB 886 --gdqfpmwntvetaing--sksedqclsqldellcpptlvegrndekalleqlvsfl 940  
 QY 1086 ALRNFDELEEDRALGIPELVSQSAVD--PBOFSSQDSN--IMLEOKAPVPPQVQASQA 1141  
 DB 941 sgkdetelelaeldralgldklv-gggglldvlserfppqgqtpplimeerplnlyqyspfl 999  
 QY 1142 QMAQGSYSPMDQPNFHTMQRSPYATLRMQPP-----GLRPTGLVON--QPNOL 1189  
 DB 1000 pta-----lpsrfqgmtrqkpslgtmrvqtpvrgafsgpmgmprqlntppaapql 1054  
 QY 1190 RLQLOHRLQAO-----QNRQPLQISNYSNVLTLRPQ-----VPTQAPINQMLAQOR 1240  
 DB 1055 rqlqgrlqgqgqllhqnrgallnqfaataqvgymrgsmgqqlcpqpplnsgmlaqrg 1114  
 QY 1241 ETLNQHRLOROMHQOQVOQRTLMRGO--GLNMTPSNVAFSGMPTMNSPRIQANAAQ 1298  
 DB 1115 elysqghrqlrlq-----qgramlmgqsfgnllps-----sglpvqtnprlpqgappq 1166  
 QY 1299 PEPFNPVGIQOQPDGFGG-----ATTPQSPV-----MSPRNA 1331  
 DB 1167 fpyppnygungpctppaspslsqilaanpeaslaanrnmvsrgmtgnlqgfgtgnlmpq 1226  
 QY 1332 HT-----QSPMMAQSOANPAYQAPSDINGM 1356  
 DB 1227 qnvfygpgagmvpqgeanlaspsssmvmpipppqssllqgfprrasgyqsp-dmkaw 1285  
 QY 1357 AOGNNGNSMFSQ--QSPRHFGQANTSMYNNNNININVMATNTGSMSSMNQTOISWT 1414  
 DB 1286 qggaglnnnvfqavqnpq--tpaqpyvy-dnmsltsvmagqnlvnygmnpmmagmqm- 1340  
 QY 1415 SVTSVTSGLSGMSPREOVNDPALRGNLFPNOLPGMDMKOGGCDTRK 1462  
 DB 1341 --sliqpmgmntlvceqindpalrntglycnqslstdllktaadgtq 1366

RESULT 8  
 ID AAM99481  
 AC AAM99481 standard; Protein: 1362 AA.  
 XX AAM99481;  
 XX 08-JUN-1999 (first entry)  
 DE Murine p/CIP protein.  
 KW Mouse; p/CIP, p300/CBP/co-integrator-associated protein; gene expression;  
 KW breast cancer; inflammatory disease; atherosclerosis; osteoporosis; ds.

OS	Mus sp.	Location/Qualifiers
XX	Key	540
XX	Misc-difference	/note= "defined in specification as ?"
FT	Misc-difference	1080
FT	Misc-difference	/note= "defined in specification as 0"
FT	Misc-difference	44..45
FT	Misc-difference	/note= "there appears to be 20 amino acids missing between these positions in the sequence given in the specification"
FT	Misc-difference	1044..1045
FT	Misc-difference	/note= "there appears to be 20 amino acids missing between these positions in the sequence given in the specification"
XX	17-DEC-1998.	
XX	12-JUN-1998;	98WO-US12263.
XX	12-JUN-1997;	97US-0049452.
XX	(RESC ) UNITV CALIFORNIA.	
XX	Glass CK, Rose DW, Rosenfield MG, Torchia J;	
XX	WPI: 1999-080883/07.	
XX	N-PSDB; MAX26000.	
XX	New nucleic acids encoding p/CIP and Ncoa-2 polypeptides - are used to identify agents that regulate gene expression, e.g. for treatment of cancer, inflammatory disease and osteoporosis	
XX	Claim 16; Fig 1; 100pp; English.	
XX	This sequence represents the amino acid sequence of the mouse p/CIP (p300)/CBP-co-integrator-associated protein) polypeptide. The protein can regulate gene expression so are potentially useful therapeutically, e.g. against (breast) cancer, inflammatory disease (e.g. atherosclerosis) or osteoporosis.	
XX	Sequence 1362 AA;	
XX	Query Match	27.9%; Score 2127; DB 20; Length 1362;
XX	Best Local Similarity	39.2%; Pred. No. 4.7e-130;
XX	Matches 543; Conservative 214; Mismatches 424; Indels 204; Gaps 44	
QY	1 MSGMGENTSDPSRAETRRKRECPDOLGSPSKRTTEKRNROENKYIELEALIFANENDI 60	
DB	1 msglgsstldplaaesrklklpcdargqvglysgekwttrgeek----- 44	
QY	61 DNEFKPRDKCALIKETVVKQIRQIKEQEKAAANIDEVQKSDVSTGGGVTDKDALGPMIL 120	
DB	45 ----vkprdkcaliketvtrqirqikeqgkttssd-dvqkadavstggvtdkdaip111 99	
QY	121 EALDGEFFVNLGEGNVFVSSENTQYLRVQOELMKNKSVYSLIHLVGHDTFVKRLPKSI 180	
DB	100 qaldgtlflvvntrignlvfyseentqylqyqkdglvntsvstlhprrkdfllntyqnpq1 159	
QY	181 VNGGSSGSEPRRNSHTFNCRLVLPKLPDSEEGHNDQEAHOKYETMQCPAVSQPSKIKE 240	
DB	160 melglgmrttdkkaayllyltmllmk-lhldledvnaaspetrqrlyetlqcfalsqprml 218	
QY	241 EGEDLOSLCLIVARVPMKERPVLPSESEFTTRDLOGKTTSDTSMRAAMKPGWEDLV 300	
DB	219 egedlqcmclvavr---tarpfsspestfltrhdlsqkvnltdnsltsmrgfdil 275	
QY	301 RRCIQFHAHQHEGESVSYAKRHNNHEVLKGLAFSQLIRFSLSDGLVLAQOTKSLIRSQT 360	
DB	276 rrciqfifsfalndqgsws-qtrhygeavvhhgaetpryrfalsldgtlvsagtksklfnrv 334	

QY	361	TNEPOLVLSLHMLHREONVVCVMMBDLTGOTMGRPLMPLISSN----	PAHOLCSGNQOD	416
Db	335	IndrIngfistftblqre-----	gyrpnrlpqdkxlrppaagcgvmsbpqn	381
QY	417	MTLSSNINFRPngKBEQMGMGRFGSGGMNVS--GMQATPR--	QGSNVALMKNSPSQ	472
Db	382	vqmgsttygprpdsntqmgagatgyassavaalrpgqslqpspsyqunsgylmspsph		441
QY	473	SSPGMNEGQPMPLSPKRRMSPGVASSPRIPEQSFSPASLSLPGVGSSTGNSHSYTN		532
Db	442	gsppqjgpnqgnlmsprn-----	gspkmasqfipaagahmspmgsgntcg-shsfess	494
QY	533	SLNALQALSEGHVSLGSSSLASPDLMKGNLQNSPVMMNPPLSKMGSLSDSKDCGELGE-		591
Db	495	slsaqltaisevgysllstlsspgk---ldnsp-nmnlsgpskvsqdxkspxlgyceq		550
QY	592	-PSEGTTCQAASSCHPEQKETMDPNLPRVSSERADGOSRLHDSKQTKLQLLTTKSD		650
Db	551	npvessvqgsnr-dbpvkkksskes-----	sgvseirprplleakgnhklqllqtcsd	603
QY	651	QMEPSPPLASSLDN-KDS-----	TGSLPGSGSTGHTSLKEKHILRLQ	695
Db	604	drghslinspldpckdxsvstpsgyssstsgvassstnvnbsllqekhrllhklq		663
QY	696	DSSSPVDLAKTAEATGCKDLSOBSSSTAPESEYTIKQEPVSPKKE-NALLRYLLDKDPT		754
Db	664	ngnspraeavaktlaetgkdd--	tstaesgegturqeqslpkkemallrlylldrddp	719
QY	755	KDIGLEPTEKLELSDKTPPASNTKLIAMKTEKE-EMSEFEGDQGSLEDMLELDDL		813
Db	720	sdvlatelqpbadqsdsklqscscsthpssgqgdklkteendevsgldldldldldl		779
QY	814	QNSQLPOLFEPPTRGAGAGASVDOQAIIINDLMQLTAENSPVTPVCAQKVALRISQTFMNP		873
Db	780	tss---dfymptctgphga-kgqmf-----	agreslglr-----sp	812
QY	874	RPGQLRLLPNQNLPLDITLQSPGAGPPPIRNSPYVIP-QPMCMNOGMIGNQNL		932
Db	813	qvsvgrppryravsld-----	spvsvsgqpvkxvsafgplkqprlqgnprmdsgny	868
QY	933	GNSTGMITGNSASRPMPSGEMA-POSSAVR-----	TCAATTSAMRPPVQGM	981
Db	869	g-amngprnrvpvnftsspgdwglansasrmepllasspigtgadysatlprpamgsv		927
QY	982	RNPASIRMPRSSOPG-----	OROTLOSUVNINISPELENNMCGP	1021
Db	928	-----	plrlsrnrlprgarpslqgqgqgqgqgqgqgqgmqrteigpmgyn	983
QY	1022	QYSQOQADPNQTAWPESILPIDQASASONRQPFSSPDLLICHPAAESPSDEGALLD		1081
Db	984	pys-pavsgnpgsgwpegmglsmegprbgsgnrlrlfnslldl-----		1024
QY	1082	QLYIALRNFDELLEIDALGILPELVSQAVDEQ--FSSQDSNIMLEBKAAVFPQOAS		1139
Db	1025	hflsnlntatgleeidalqaprlvngqgaleekqdvfgqgeaavmmdkaalyxtypa		1084
QY	1140	QAMQMGYSYPM--DDPNPHYH-----	GGRPYATLRMQPRGL-RPGLVQONPNOLRLQ	1193
Db	1085	gpprlqggfmlqgqpsfnsnmqglsgqgsfplqgmhpragrlvtrp--	tnbrpkrlmqll	1141
QY	1194	QHRLOAQ-----	QNRQPLMNOJNSNVSNNULRAGVYTOAPINAMLOARORELIMQHLRO	1249
Db	1142	qqrllqgqgflngsvrqalemkmpag-tawmrmpm-qgafinaqaagqkrelmsnhlqg		1199
QY	1250	ROMHQOQOVQORTLMNR--	GQGLNMTPSWAVASGMPATMSNBRIFQANAQOFPPRYNG	1306
Db	1200	qrm-----	ammusypyqgafsprrnvtlaasmdyLaagsampqapppqfpypany	1250
QY	1307	ISQOP	1311	
Db	1251	tgppp	1255	



```

RESULT 9
AAW81028
ID AAW81028 standard; Protein; 951 AA.
XX
AC AAW81028;
XX
DT 12-MAY-1999 (first entry)
XX
DE ER Interacting domain of AIB1 protein.
XX
KW AIB1; amplified in breast cancer; cancer; steroid; receptor;
KW coactivator; SCR; estrogen; ER; estrogen dependent transcription;
KW breast cancer; lung cancer; colon cancer; prostate cancer;
KW melanoma; PAS region.
XX
OS Homo sapiens.
XX
PN W09857982-A2.
XX
PD 23-DEC-1998.
XX
PF 17-JUN-1998; 98WO-US12689.
XX
PR 17-JUN-1997; 97US-0049728.
XX
PA (USGO ) US GOVERNMENT.
XX
PI Meltzer P, Trent JM.
XX
DR WPI: 1999-080946/07.
XX
PT New isolated steroid receptor co-activator, AIB1 - used to develop
PT products for the diagnosis and treatment of steroid-responsive
PT tumours, e.g. breast, lung, prostate or colon cancers or melanomas
XX
PS Claim 6; Page 34-35; 57pp; English.
XX
CC The AIB1 protein is a member of the steroid receptor coactivator-1
CC (SRC-1) family of nuclear receptor co-activators that interact with
CC estrogen receptors (ER) to enhance ER-dependent transcription.
CC The AIB1 gene is amplified and over-expressed in certain cancers
CC in particular breast cancer and steroid hormone responsive cancers.
CC The AIB1 polypeptide can be used to identify compounds which inhibit
CC ER-dependent transcription. Increased expression of the AIB1 gene
CC indicates aberrantly proliferating cells, thus detection of
CC increased expression of the AIB1 gene or an increase in the number
CC of copies of the AIB1 gene can be used to diagnose cancer or a
CC predisposition towards developing cancer. Compounds which inhibit
CC expression of AIB1 or compounds which inhibit interaction of AIB1
CC with steroid receptors or nuclear co-factors can be used for
CC reducing the proliferation of cancer cells. The PAS domain of the
CC AIB protein is a highly conserved domain among the SRC-1 family of
CC proteins. The ER-interacting domain of AIB1 mediates binding of
CC AIB1 with a steroid hormone receptor protein.
XX
SQ Sequence 951 AA;

Query Match 19.5%; Score 1486.5; DB 20; Length 951;
Best Local Similarity 38.4%; Pred. No. 2.3e-88;
Matches 392; Conservative 164; Mismatches 332; Indels 133; Gaps 37;

OY 303 CIOKFAHQEGESVSUARHHHEVLRQGLARFQSLRFSLSQDTLVLAQTKSKLINSQTTN 362
DB 1 CIGFTFSIndqsws-qktrhygeaylInghaeLpvyrfIsldqtlvtbaQtkSLfInpYtn 59
OY 363 EPOLVISHLHMRQENVCMNPDLGOTMGRPLNPISSNSPAAHQALCSGNPCQDUTLSSN 422
DB 60 drhgfvshfIqregnyrrpnpvggIirpImagcnsVvg----mmsmpqgIqmpss 115
OY 423 INPFGPKRQMGMPGRFGSGGNNHVS---GMOA-TTPQGSNYALKMNSPQSSPCMN 478
DB 423 INPFGPKRQMGMPGRFGSGGNNHVS---GMOA-TTPQGSNYALKMNSPQSSPCMN 478

```

```

DB 116 raygladpSttggmsgarygssnsIasItpqpmqgssyqnnnygInmspphsgpyla 175
OY 479 PGQPTMLSPRIHMSPGVAGSPRIIPSOFSFAGSLHPYGVCSSTGNSHSYNSLNAIQ 538
DB 176 pmqnamIsprnr-----gspxIashqfIsPvagvshpmassgntgn-hsfssssIsaIq 228
OY 539 ALSEGHVSLSGSSLPDLKMGNLQNSPVNMNPPRLSKMGSLDSKXCFGLYGF--PSECF 596
DB 229 aIsgevvtIsIstIsSpqk--lOnsp-nmItqpskvnsqskspIgfycdqpnvess 284
OY 597 TGOAESCHPGEQKETNDNPLPAVSSERADGQSRILHDSKQGTQLLQTTKSD-----Q 651
DB 285 mcqsnrdhIsd-keskes-----svegaenqrgrpleekykhkIlqItcsddtghss 337
OY 652 MEPSPLASL-----SDTNKDSGLSPGSGSTHGTSLKEKHILHRLLDSSSPV 701
DB 338 ltnspIdssckessvstpspsvstsgvssstsmshgslIqekhrIlhklIngnspa 397
OY 702 DLAKLFAEATGKDLQSESSSTAAGSEVTIKOEVPSPKKE-NALLRLYLKDQDTRIGLP 760
DB 398 evakItaeatgyd-----tsItscgdgnvvkqeqIspkKkenallIyIlIdrdpsdaIsk 454
OY 761 EITPKLERLDSKTPASNTKLIAMKTEKE-EMSFEPGDPGSELDNLEELDLQNSQLP 819
DB 455 eIqpvyevydnkmsqctstIspsgekdpxIkItseesgdlndIdaIdItssdfy 514
OY 820 QLEPDRPCAPAGSVDKAIIINDLMQLTAEBSVTPVGAOKTALRISQSTFNNPRFQLG 879
DB 515 nnsIsen-gshIgt-kqgvfQ-----gnsIqIksqsv-gslrp----- 551
OY 880 RLLPNQNLPLD--ITLQSTGAGPPPIRINSSPYSLP--OPGMGMOGIGINGNSS 936
DB 552 -----pynravsIdspvsvsppvXnIsaIfmIpIpkqpmIggnprrmIdsgenyssm 603
OY 937 TGMIGNSASRPMPSPGEMA-POSSAVRYTCAATTS-----AMNRPVQGMIRNPA 985
DB 604 ggpnrnrvltqIppsgdwlprnskagImemnsnmgrrpgdyntslprpalgsI----- 659
OY 986 ASIFMRPSSQPOROQL--QSOVMNIGSELEMNMGFPQYSQOQAPRNOTAPWESILPI 1043
DB 660 pElprIsnspIarpyIqgqgmIqmrpgeIpmnganpyg--gaasngIgwawpdmIsm 718
OY 1044 DOASFASQOROPFPDCLCRPHRAESPSDEGALLDQLYALRNF--GHEELDRALG 1101
DB 719 egyptqtrpIlnsIddlvgrpsnlEegsderalIdqIhclIsntdaqtLeIdralg 778
OY 1102 IPELVSOQAVDPEQ--FSSQDSNIMLEQKAVFPQOYSAQOAMQGSYSPMQDPNFFHTM 1159
DB 779 IpeIvngqgaIepkqdaIgggaavmmqdkagIyqqtYpaqgppmg-----gfhlq 830
OY 1160 GORPSTATLRMO-----PRGELRPTGLVQ---NOPQLRIQLQHLQAO-----QNRQ 1204
DB 831 gqspsfmsnmnmqmgqnfIrlqgmIpranImrptnlrqlrImqIqqrIqgqfInsgrq 890
OY 1205 PLMNQISN-VSNVNLILRPGV--PQAPINAOQLARQREIILNQLRQR--MHQOQOYQ 1259
DB 891 aLeIkmenPtagaavvmrImqpqgfIlnaQvagrstElshlrItrvvaammngqgqgq 950
OY 1260 Q 1260
DB 951 q 951

RESULT 10
AAW26370
ID AAW26370 standard; Protein; 1061 AA.
XX
AC AAW26370;
XX
DT 04-DEC-1997 (first entry)
XX
DE Human steroid receptor coactivator-1 (SRC-1).
XX

```





OY	338	SQTNNEPOLVLSHMLHREON-----VCYMNDLT-----	387
Db	418	nltpbgedfimsyvtllnsendmnsntcgaagglgldglcamapspelaisslsmdd	477
OY	388	-----GQTMGKRLNLISSNS-PAH--CALCSGNP-----GOQMTLSSNINP-----INCP-	430
Db	478	glnbgtygssassspaaagmlphtlllgldygqgqgqggnstctctvgpblmsaaltgtg	537
OY	431	KEOMGMPWGRFGSGSGNNH-----VSGMOAT-----TPQGSNYALK-	466
Db	538	lqqqqqqqrgsagaassassassanaalvafasapaeahsfygsdtfeidiahsfeldp	597
OY	467	---MNSFSQSSPGCNPNQPSMLSPR-----HRMSPEVACSPIRPSQSS-PAGSLHSYVG	518
Db	598	sgvgawcdsrpmsraasvafvtpirpssghgfspavcaspatyqlsishaaslipspqs	657
OY	519	VCCSSGNHSHSTNSLSMALALBEGHGVSLGSSLASPDLKMNLNOSPVMNMPPLSLKMG	578
Db	658	nasagggayvgfntfh-----sdsadvkpeadvqqnqqsnm-----n	697
OY	579	SLDSKDEFGVLGEFSEGGTGOAESCHPCBOKETENDPNLPPAVSSEBARDGOSRLHDSKGQ	638
Db	698	ssasnpllg--gglpnyvggmlls--qqqgqqgt-----ppqgqqqqqqes-----	740
OY	639	TKLLQLLTSSKDOMESPRLASSL-----SPTNKDSITGSLGSGSTHGTSL-	683
Db	741	serlrlhlltkagm-----agglgldgdekyfkpegseeekhaasgfkmgqpgqgmuf	795
OY	684	-----KKKHILRLRLQDSSPPLDLAKLAEATGTGDLQESSSTAPGCEVITKCEP	734
Db	796	gpmgsmgrgyvgnssmllk--agnsqnbn--llkllnks--eddongsgggpgs-----	844
OY	735	VSPKKENALLRYLLDDK-----DTKDIGLPEITPKLERLDSKTDPAASNT	779
Db	845	-----mmnsrgelmrqtknpdgsgshgmhrnsaagmstedl-----kamklqsdpslnr	895
OY	780	KLIAMKTEKEMSEPEPDQGSSELNLELIDLQNSQLPQLPFDTRCARPAGSVDKAI	839
Db	896	k-----rslneppddpsakrs-----edkp-----sklct	920
OY	840	INDLMQLAENSPTVPYGAQKALTALRISQSTFNPNREGCLLPNOMNPLDITLQSPIGA	899
Db	921	gqkmlakllgnppklypa-----pnpeqplqyk-----	948
OY	900	GPEPPIRNSPEYVIPQPGMNGNOIGNOGLNSTTGMIGNSARPTMPSGEWAPOSS	959
Db	949	-tlpdlstsvstslaapgnllsag-----stg-----pxaa	979
OY	960	AVRYTCAATTSAMNRPPVOGGMIRNPAASTPMRSSOPGORLQLOSVMNIGSELEMMNG	1019
Db	980	anr-----nrkqgq-----qqgqqqqqqqqqqvagi--psq-----	1008
OY	1020	GPOYSOOQAPRPNQAPBESILPDIQASFSQNRORFGSSPDLCLCPHPAESPDEGAL	1079
Db	1009	-----qqqqqpnndyilsqqqqqqqlqspbqlafqngqlattatstislttaastaaaaaai	1063
OY	1080	LDOLYTLALRNFDGLEIADRALG-----IPELYSOSQAVDPEQFSSQDSNIMLE	1127
Db	1064	lg-----egseslskllsdwemeyppddprlivenapasesaalndiq-----kslmd	1109
OY	1128	OKAPRFPOQVASQOMAMAGSTSPKQDPNFTTMCQRPSTVTLMKQRPRLGRLPGVLVO--NQ	1185
Db	1110	vesaafigndlmgltmtyq-----qqghqgqqgqqglalqlaqgqqqqrghlqppay	1163
OY	1186	PNQLRL-OLQHRLOAQONRQRPIMNQISNVSN-----VNLRLRG-	1223
Db	1164	pgmllmqqqqghqqgqngqqlmgtrleamngnggfqrppmnypargrgpmnavafrpg	1223
OY	1224	---VPTQAI-----NAQMLAORGEILNQLRORONHOQVOQFLLMNGKGLMWTSPM	1276
Db	1224	vvlpaqgqllnrlqqqlaaaqke-----tllqqqgqqllrvpenatgma--glnlgsi	1278
OY	1277	VAPSGMP-ATMSNRIPOANAQOFPEPPNYG-----ISQADPGFTGATTPQ-----	1322

```

Db      1279  IntGpavrsIsrtnlp-sdaq-----IspnIaqlmqqIsprIsapysqpnpqyapfp 1334
Qy      1323  -----SPLMSPRAHfHQSPPM00SQ-----ANPAYQAPSDING-- 1355
Db      1355  qpqrIIsppqqqqlsqqqnnvqqqqqIayqqqvvgdgrsntprfgsnsmgspgmqnsbpq 1394
Qy      1356  -WAGNMKGNS-----MFSQSPSPRHGQOANTSMY-----SNMNNINTVMSA 1395
Db      1395  qwsgsggggggpgpplpsgnaagrtIqhnmlIaqlqvgvpynarqyqnnqrrtIpspa 1454
Qy      1396  TINTG-----MSNMNQMTGQISMTSVTSVTSRSGLSMGPEQVNDPALRGCNLEPNQ 1446
Db      1455  vgpqnpaaqaalqrrqnsfgqg-gggagatIpdsgsvvgfsgppq---spyglnvnlfgqq 1508

```

## RESULT 13

ID AAW78487 standard; Protein; 399 AA.

AC AA78487;

DT 21-MAY-1999 (first entry)

Human SRC-1 ligand binding region.

KW Human; SRC-1; ligand binding; fusion protein; screening;

KW intranuclear receptor protein.

OS Homo sapiens.

PN JP11032767-A.

PD 09-FEB-1999.

PF 16-JUL-1997; 97JP-0191007.

PR 16-JUL-1997; 97JP-0191007.

PA (YAMA ) YAMANOUCHI PHARM CO LTD.

DR WPI: 1999-183824/16.

XX

PT intranuclear receptor-combining protein

PS Claim 3; Page 10-11; 23pp; Japanese

CC This sequence represents the human SRC-1 protein ligand binding region

CC protein can be fused to reporter proteins and used

[illegible]

SQ Sequence 399 AA;

Query Match 5.28; Score 397; DB 20; Length 399;

Matches 142; Conservative 70; Mismatches 139; Indels 84; Gaps 23;

QY 395 LNPISNSPAHQALCSGN--PGQDMTSSNINFPINGPKEQ-----MGMPMGRF 441

```

Db 7 vnp--sispahqvarsstlppsnsmvstrln-----rqssdlhssshsnssnsqsf 58

```

442 GGGGMNHVSGM-----OATTPGGSNYALKMNSPSSSPGMPGOTSMSPRHRMSPGV 496

59 acspasaiyanva]nkaass-asksps]n]nnpmeatais]aq---fmsprvysq] 114

OV A97 AGSPRTDPSQFSD-AGST.HSPVGVCS--TGNSHSYTNSSINATQATSEGHGVST.GSSIA 553

[illegible]

[illegible]

RESULT 14  
 ABB65772  
 ID ABB65772 standard; Protein; 5533 AA.  
 AC  
 XX ABB65772:  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster polypeptide SEQ ID NO 24108.  
 XX  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.  
 XX  
 XX Drosophila melanogaster.  
 OS  
 PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PE 23-MAR-2001; 2001WO-US09231.  
 XX  
 XX 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 XX (PEKE ) PE CORP NY.  
 PA  
 XX  
 PI Venter JC, Adams M, Li PWD, Myers EW;  
 XX  
 DR WPI; 2001-656860/75.  
 DR N-PSDB; ABL09875.  
 XX  
 XX New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 PT  
 XX  
 PS Disclosure; SEQ ID NO 24108; 21pp + Sequence Listing; English.  
 XX  
 XX The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL130511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins  
 CC (ABB57737-ABB72072).  
 CC  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 XX Sequence 5533 AA:  
 XX

Query Match	4.4%	Score 334	DB 22	Length 5533
Best Local Similarity	20.3%	Pred. No. 1e-11		
Matches 331	Conservative 197	Mismatches 654	Indels 450	Gaps 74
QY	8	TDDPSAEERKKRECECDOLGSPRKTERKRNDEQEKYIEELAEELIFANFNIDNFNKP	67	
Db	2819	tsapstlaqsktrg-----	2851	
QY	68	DKCALKEFVVKQIROLKEQEKAAANIDEVOKSDVSTSGOYIDKDALPMMLEALDGF	127	
Db	2852	ega-----enlk-----dlpsvst-----gvlvaalg-----	2877	
QY	128	FVYNIEGVNVFSENVTOYLRNOEELMKSVYSILHV--GDHTEFVKMLPRKSIYNGS	185	
Db	2878	-----tykqesprtskneehidqltvhepeqqqlertslrsgs	2917	
QY	186	MSGEPRRNSHFNCRMLYKRLPDSSEEBHNOEHAQKLETMOCAVASPKIT--KEEE	243	
Db	2918	ssashaderhrrekr--ekkrkrsqreqnq-lhqskskvet-kvddadnsvdmdaegr	2973	
QY	244	DLOSICIVARRPMKREPLVPSSEFTTRDLOGITGLDSTYRAAKPMGMDLVRRC	303	
Db	2974	aleaql-----msdftrkise-----earpstaatyrsmdt-----	3006	
QY	304	IOKFNAOHGESVSYAKRNHNEVLROGLAFSOLYRPSLSDGTLYVAOTKSLIRSQTWE	363	
Db	3007	virfenehmsv-----dmktgyvseqeqhkskd-----kkskskskeekq	3051	
QY	364	POLVTSILHLHREOVNCPNPLDTCGTOMK-----PLNPISSSPAHQALC	409	
Db	3052	ekl---lqqqrteslprvastesaprtppqklvtnyqaasakhadqlidakhlisps---pvc	3105	
QY	410	SGNPQGDMTLSNINPPIINGRKEOGMGMPCRGSGGGMHVSQMATTPOGSNVALKMS	469	
Db	3106	kpspslpcrligdddaalhtprakrtpsr--gnqgl-----	3142	
QY	470	PSQSPGMNPGOPTMSLSPRHMSPGVAGSPRIPOFSFAGSLSPVGVCSGTNSHSY	529	
Db	3143	ptrekp-----llisp-----lptkrlpianstltsqsaetpv-----sgtvl-	3182	
QY	530	TMSLNLALDSEGHGVSIGSSLASBDLKMGLNPSVNMNPPRLSKMSGLDSKCF--G	587	
Db	3183	-ssalatpftslaagvs-----aap-----glhnspt-----sasagckkksflpg	3224	
QY	588	LVGESEEGTGAQESSCHRGEOKEF-----NDNLPRAVSSERADGSRLLHDSKGQTK	640	
Db	3225	fdgjdldrtiseaavgi-saefnstsldlnadekhlipraspratkrldkleeksrvt	3283	
QY	641	LIQLLTTSKDOMEPPLASSLSDTKNDSTGSLPGSGSTHGHTLKKRKHILHRLDSSSP	700	
Db	3284	isq-----eetesavaallgesfgetstcdysldgmdemsvneletprl---vleapdee	3336	
QY	701	VDLAKLTAFAATKDIS-----QESSITAPGSEVTIKQEPV-----SPKKENALLRY	747	
Db	3337	aalaakataetagepasllieepemeperaeapdpdeaaiesepvvevldpeelnkav--q	3394	
QY	748	LIDKDDTKOIGRLPITPKLER--LDSKTPDASNNKLAMKEKEMSPREDDOGSELD	804	
Db	3395	slkneemmdikad--tpgsertldqldtdeenpdeadsgrpslkldetvqsssspeksts	3452	
QY	805	NLEETLIDLQNSQLPOLPFDTRPGAPAGSVDRQQAIIINDLMOTLAENSPTV--PVGAKQKAL	863	
Db	3453	msnpptreanldipnv--eegpklsmestpbspvrltkrlpl--dtrktpyagllpppav	3507	
QY	864	KLSQSTFNN-----PRGGQIGR-LILPOMTLPD-----	890	
Db	3508	kleppltsklqqlvpgvqtlvlpabstfsgjsaansvlnldsnvisscntsaasatas	3567	
QY	891	-----TTLQSPGAGFPPIRINSPPSVLPQPMGMOGICIGOGNLGSSSTCMGNASR	946	
Db	3568	asasisfsgpsas-----qamqpascpkqpllcpqgalitqslimpprltspdeqtrph	3621	
QY	947	PMPGGEVNAPOSSAVRYTCAATTSAMNRPVQGM-----IRNPAISIPMRPSQOPOROTL	1002	



Db 3284 isg-----eetasavaaligesfgtsctdyldgmdemsvneletptl---vinaepdee 3336  
 QY 701 VDLAKLTAATGKDL-----QESSSTAPGEVYTIKOEYV-----SPKKKNALLRY 747  
 Db 3337 aalaakateagepaslleepemereaeapdpdeeelesepvevldpdelkav--q 3334  
 QY 748 LLDKDKTDIGLPEITFKLER---LDSKTDPASTKLIAMKTEKEMSFEPGDPGSELD 804  
 Db 3395 slkheadmmdlkad--tpgsectldtdcteenpdeasgsplkldetvgsasspeksis 3452  
 QY 805 NLEETLDDLONSLPOLFPDTRPGAPAGSVDKAIIINDLMOLTRENSPVY--PVAQAKTAL 863  
 Db 3453 nnsptpretanidipnv--esqpklsneslcpqpsvicklplf---dlcpktpvaglppspv 3507  
 QY 864 RISGSTFNN-----PRPGQLGR-LLPNQLPLD----- 890  
 Db 3508 kileptlisklqgplvqpvqtlpaphstgqisavsvlnldlnvlsccntsasaatas 3567  
 QY 891 -----ITLQSPGTGAPFPPIRNSPFYVLPFGMANGNOGMIGNOLNNSSTGMTGNSASR 946  
 Db 3568 asasisfgsptlas-----qnamqgastpkqgpltpqgalrtqslmqpplispeqtph 3621  
 QY 947 PTMPSGEMAPQSSAIVRTCAATTSAMNRPVOGK-----IRNPAASIPMRSSQPGQKOTL 1002  
 Db 3622 fawpqnvlspqs-----hhpqgplqymvgzraps---phspilhsqgrgvaq 3664  
 QY 1003 OSQVMNIGSPSELENNMGPOYS-----OOQAPPNQ-----TAPPESTLTPIDQAS--FAS 1050  
 Db 3665 srlvgqisp-----vgipmvsqpspqgqvtqgqhallltspsnsilspaspltrvlis 3718  
 QY 1051 QNRQPFCS-----SPDLLLCP-HPAESPSEDEGALLDQLYLALNPFGLIEIDALGIPE 1104  
 Db 3719 ssnspstksnysyprnqgppqpsksvaevqtlmtcltpikmtbipqyphn-----pt 3774  
 QY 1105 LVSGQSAVDEPQESSQD-----SNML-----EQKAPVPPQOYASQAQMAQG 1146  
 Db 3775 liskvlevtpqgaltqgvasssprlgsjlpbhkvnlhahbqgqgqvialakmtahbqgmq 3834  
 QY 1147 SY-----SPMODNFHMGORPSVATLRMQPRGLARTGLVONP---NQLRL--- 1191  
 Db 3835 qfmhqmhlgrqgmhmqgqllhgssqlltsa---pqhmqhqbhaqqgqghnqhnlmg 3889  
 QY 1192 -----QLOHLQAQONROPLMNQISNVSNNLTLRPGVTPQAPINAGMLAQOREIL 1243  
 Db 3890 lhaqgthctqkqhgqgqfnglqhgsgq-----qhmqgqgqgqghlsq 3935  
 QY 1244 NQHLRQRMHQOOVOQRTL-----MMRG-QGLNMTPSMVAAPSG-----MPATMSNPR 1290  
 Db 3936 qghbsqgqlnqbhaqqgqlqgkllqgmhbpqgqkspgqvghllygsstisfasqghnsq 3995  
 QY 1291 IP-----QANAOOF-----PPPPNYGIS-----QOPDQGTGATTPQSPPLMSPRMAHTQS 1335  
 Db 3996 lpargvppqgqplshspckpntllvsngvgppallltvgsbsqngqgqllphqgs 4055  
 QY 1336 -----PMOQSOQ-----ANPAYQAPSDINGMAGNMGSMSQSPHFQGOAATSMYSNNMN 1389  
 Db 4056 sgphbqgqlspsganlrlqplnvlqntpkllvgqhvaqngvpppqgqghalhyppqgq 4115  
 QY 1390 INVSMAINTGMSMNMOTGQISMTSV-----TSVTSGLSSMGPEQVNDPLRGNGLFPN 1445  
 Db 4116 kd-----stppghvpeptpsasqktsesvsvlrlptclglavisaantvgs--llteenlikl 4171  
 QY 1446 QLPQMD-MIKOE 1456  
 Db 4172 sqpkqdelieqd 4183

DT 26-MAR-2002 (first entry)  
 DE Drosophila melanogaster polypeptide SEQ ID NO 11742.  
 XX  
 KW Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN W0200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US09231.  
 XX  
 PR 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li PWD, Myers EW;  
 XX  
 DR WPI; 2001-656860/75.  
 DR N-PDB; ABL05753.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -  
 PT  
 PS Disclosure; SEQ ID NO 11742; 21pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL10511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABR57737-ABR72072).  
 CC  
 CC The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pcl\_sequences.  
 CC  
 XX Sequence 2280 AA;  
 XX  
 Query Match 4.48; Score 332; DB 22; Length 2280;  
 Best Local Similarity 20.68; Pred. No. 3.5e-12;  
 Matches 266; Conservative 138; Mismatches 452; Indels 438; Gaps 59;  
 QY 426 PINPKEDMGMPGRFGSGGMNHS-----GMQATTPOG-----SNY 463  
 Db 1017 plngdaad-gasvvgiggaagvmmvnaaglpisagqglatqpsllnprvtglvqistk 1075  
 QY 464 ALKMNPSQSSPGM--NPGQPTSM-----LSPRHRMSPGVAG-----SP 500  
 Db 1076 kctiaspalsprvlllepeqdcskkykhanglryhshabhaaggsamedsmqapedp 1135  
 QY 501 RIPPQSPAGSLSPVGVCSSTGNSHYTNSLALQALSEGKGVSLGSLASPDLKMG 560  
 Db 1136 atppsp-----gvasgtlsgasvasavpa--lapaagq---gtvaavspnplla 1179  
 QY 561 NLONSVPVNMN-----PPPLSKMGSLSKDCFCLY 589  
 Db 1180 nsnprvtngnavapsapatgsvtiaapntpsvvetqaplltpppyt----- 1225  
 QY 590 GEPEEGTTGAESSCHPEQKETND---PNLPPAVSSERADGQRLHDSKGQTKLL-- 642  
 Db 1226 --ppaprplicavat--pgaegsvsvlplgnlpltagpnastqgpprtqgqqllyvpq 1281  
 QY 643 -QLUTTKSDQMEPSPPLA-----SSLSDTKKSDTGSILPGSGSTHGTSLKAKKHLH 691  
 Db 1282 gsaaslgqgqgqgqpvavagssilagslsgalsqhgqimvgilp-----amlsdqgq--q 1332

RESULT 16  
 ABB61650  
 ID ABB61650 standard; Protein; 2280 AA.  
 XX  
 AC ABB61650;  
 XX





QY 303 CIOGFHAGESESVYAKRHHHEVLROGLAFSQRFR----- 339  
 Db 353 sl--fscqhgads-erlmatkfsvlskqggetstryrlfigkygycwllsqatlvdyklkr 409  
 QY 340 -----SLSDGTLVAOTKSKLINSQTNFQOLVLSLHMLHREONV 379  
 Db 410 qsvvcynvynslnknkheiyislaqtaaseqkqhgaatekepekaadellaqetke 469  
 QY 380 CVMNPDLTGOTMGRPLN-----PISSNPAHQ--ALCSGNPGQDM 417  
 Db 470 tvnprhntselqakrplqdesekaektleetktiatiprvtatstaadqikqlpesnpykqi 529  
 QY 418 TLSSNINFPINGPREQMGMPMGREGSGGMNHYSGMOATTPQSGSNYALKMNSPQSSPFGM 477  
 Db 530 l-----gaellikr-----ehspgprtilaq-----llegssag 560  
 QY 478 NPGQ--PTSMLSPRHRHSPGAGSPRIPPSQSPAGSL--HSPGVCSSTGNSHSYTSS 533  
 Db 561 rpeekrkpsvavslrpsp---applrpp---Plavickktrpiv----- 599  
 QY 534 LNALQALSEGHVSLGSLASPDLMKGNLQNSPVNMNPPILSKKGSIDSKDCFLYGEPS 593  
 Db 600 -----epnl-----pptataais----- 615  
 QY 594 EGTGQAESCHPEQKFTNDPNLPRAYSSERADQSRU--HDSKQOTKLQTLTTKSDQ 651  
 Db 616 -----ssnqqlaqgqqlqppqp---aqdmksyfcslfaddgrglmlke----- 659  
 QY 652 MEPSPLASSLSDTN-----KDTGSLPGSGSTHGTSLKKEKHLMLDODSSSPV 701  
 Db 660 -epddlsbhlastrnciqldemtprfsmhvglmgt-----cllpedhsl 702  
 QY 702 DIAKTAETAGTKDLSQESS--TAP-----GSEVTRIKQEPVSPKKENALLYLKDQDTK 755  
 Db 703 dttcstaagqhqpspsstsspsntssnysanspisp----- 744  
 QY 756 DIGLEITPKLERLSDTDPASNTKLIAKTEKEMSFEPDQPSSELDLEITLDLON 815  
 Db 745 -----ltp-----nataasn-----pshqgqgqhnoqg-----qg 771  
 QY 816 SOLPQLPDDTRPGAPACSVKQAIINDLMQLTANSPTVPYGAOKTALRLSQSTFNNRP 875  
 Db 772 qqqqgqhnpqhdnsnss-----nldplfnyeesndt-----scsqhlhpsstlskspe 822  
 QY 876 GOLGRLLPNQNLPL--DITLOSPTGAGPFPIRNSSPYSVPOGMNGMGNQMLG 933  
 Db 823 ssrlpslspnslqgedfssfeaamrapiyldiddmp--litedclm----- 867  
 QY 934 NSSTGMIGNSASRPTMPSGEMAPQSSAVRYTCAATTSAMNRPVOGMI RNPAASIPMRPS 993  
 Db 868 -----wcpedlqtmv----- 878  
 QY 994 SQRGQRTLOSQVWNITPSELMMMGPOYSQOQAP-----NOTAPWESIL----- 1041  
 Db 879 --pkeldalqgqlqgqhqhgyagnltyqgqgqgqqlqghfsmalscpastvasls 936  
 QY 1042 --PIDQASFAQONQPRGSSPDD--LLCPHAPAESPDGAL----- 1079  
 Db 937 pspvqg--hmqgqgaavfssdselaallic-----gsngtllslilgsgvtyaeeener 988  
 QY 1080 -----LDOLYLALNFDLEIDRALGIPELVSQAVDPEQSSQDSNIMLEOKAVFP 1134  
 Db 989 lqghqgqgqgqtsgnefttfgqlqgelqgeeqqrgqqgqgqgqqlslisnieck 1048  
 QY 1135 QOYASQAMQAGSISPMQDNNFHTMGORPSTATLRMQPREGLRPTGLVONO-----P 1186  
 Db 1049 eky--dvymgmslchpmed-----afendyskdsanldcwdlilqmqvdtpevsp 1096  
 QY 1187 N-----QLRQLDHLRLQAOQNRPILMNQISNVNLTLEPGVPTQAPINAO 1234  
 Db 1097 naaaptrckvsaql-lqgqgqqlqgqgqgq-----qnlilnaavp-----li 1136  
 QY 1235 LAOROREILNQLROROMHOQOQVOQRTL--MMRGQGL--NMTPSNVAPSGMPATMSNPR 1290

Db 1137 tiqnkkelmqgqgqgqgqgqqlqgpaikllngaslapvntkatiirlveskppttqsr 1196  
 QY 1291 IPQANAOQFPPPPYKYSQODPG-----FTGATTPQSPILMSPRM-----AHTQSPMM-- 1338  
 Db 1197 makvnl-----YPPgqgqhknkrlhnsatqagrpveskrlksqtlcldvgsqqlq 1246  
 QY 1339 -----QOSQANPAYQAPSDINGMAQONMGNSMFQSOQSPHRGQOQANTSMYSNNMNIN 1391  
 Db 1247 qlqkdpeqgqqtgaakagse--rw-----qlsaesk-----qqkqgqgsnsvlkn 1291  
 QY 1392 VSMATNTGMSMNMQMTQGISMTSVTSVTSGL-----SSMGPPQVNDPALRGCNLF 1443  
 Db 1292 llvsgrdddsseamliednaslvqplpkygylplhchtsssvlrdymhnpilisgtfnf- 1350  
 QY 1444 PNQLP-----GMDMIKQEGDT 1459  
 Db 1351 --qlspvfygsdssgqgqdet 1368  
 RESULT 18  
 AAW79533  
 ID AAW79533 standard; Protein; 846 AA.  
 AC AAW79533;  
 XX  
 DT 02-FEB-1999 (first entry)  
 DE Human CLOCK protein.  
 KW CLOCK; circadian rhythm; human; jet-lag; sleep-wake disorder;  
 KW seasonal affective disorder; cancer; transcription factor;  
 KW diagnosis; therapy.  
 OS Homo sapiens.  
 XX  
 FH Key  
 FT Protein  
 FT /note= "Location/Qualifiers  
 FT Protein 2..846  
 FT /note= "Claimed fragment"  
 FT Protein 10..846  
 FT /note= "Claimed fragment"  
 FT Protein 11..846  
 FT /note= "Claimed fragment"  
 FT Domain 35..81  
 FT /label= bHLH  
 FT /note= "basic helix-loop helix domain"  
 FT Domain 115..163  
 FT /label= PAS-A  
 FT Domain 273..318  
 FT /label= PAS-B  
 FT  
 PN W09840514-A1.  
 PD 17-SEP-1998.  
 XX  
 PF 13-MAR-1998; 98WO-US05114.  
 XX  
 PR 30-JUN-1997; 97US-0885291.  
 PR 13-MAR-1997; 97US-0816693.  
 XX  
 PA (NOUN ) UNIV NORTHWESTERN.  
 XX  
 PI Pinto LH, Takahashi JS, Turek F;  
 XX  
 DR WPI; 1998-520828/44.  
 DR N-PSDB; AAV61450.  
 XX  
 PT Novel mouse and human circadian rhythm gene, clock - useful for  
 PT treating e.g. jet-lag, sleep-wake disorders, abnormal cell division,  
 PT etc  
 PS Claim 20; Fig 14.1-14.3; 154pp; English.  
 XX



```

OY 31 KRNETKRRNEOEKNYIEELAEILFANFNIDNFNKPKCALIKETVQIROIKOEKAA 90
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 39 rnksekkrtrdgnvnlkelsgmipgnar-----kmdkstvlqtsidflr--khkelta 89

OY 91 AANIDEVOKSDVSTGOGVIDKDALGPMLEALDGEFFVNVLEGNVFESENVTOYLRYN 150
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 90 qsdasel-rqdwkpc---flsneeftqlmlealdgfflaimgdtsilyseesvtallehl 145

OY 151 QEELMNKSVYSILHVGDTHEFVKNLLPKSIYNGSGMSEPPR-RNSHTFNCML----- 203
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 146 psdlvdgsifnflpegehsevyk-llstchllsdsllpeylyksknqlfecchmlrgtldp 204

OY 204 -----VKPLPDESEEGHND-----QEAHQ-KYETMOCFA-----VSOPKSIK 239
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 205 kepstleyvklfgnfxkslnsvssahngfegltqrthrpseydrctvatvrlapqfllk 264

OY 240 EBGEDLOSCLICVARRVPMKERPVLPSSSEFTTRDLOGKITSLDTSTMRAMKRGWEDL 299
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 265 e-----mctv-----eep-----neeftshlslewkflfld---hrappllylplf 302

OY 300 VRRCIOKFAHQEGESVSYAKRHHHEVLROGLAFSOIYRFLSDGTIVAQTKSKLIRSQ 359
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 303 evlgtsgdydyhvddlenlakchen-lmqyqkscyyrfltkgqgwawlqchyllyhq 361

OY 360 TTNEPOLVLSLHMLHREONVCMNPDLTGOTWKGKPLNPISNSPAHQALCSGNPGODMTL 419
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 362 wnspefivchtvsvsyaevraerrelg-----ieeslpetaadksqdsadnri 412

OY 420 SSNINFPINGPKQMGMPGRFGSGGNHNVSGMOATTPOGSNVALKMN--PSOSSPGMN 478
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 413 nt-----vsjkealerfdhs-----pfpasasrsrkshtavsdpst 451

OY 479 PGO-PTMSLSPRHNRHSPVAGSPRTPPSQFSPAGSLHPVGCSTGSHSTNSLNL 537
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 452 ptklptdst-----pprqnlpaa---hekml-----vqtrssfsqgslns- 487

OY 538 QALSEGHGVSLSGSLASPDLMKGNLQNSPVNMNPPLS-----KMGSLDS-KDQFGLYG 590
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 488 -----qsvgsaltcp--vmsqatnlpj---pqgmsqfqsaglgamqlkd----- 528

OY 591 EPSECTGQAESSCHPGQOKETNDPNLPRAVSSERADG-QSRLHDSK-----GQTKLQL 644
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 529 -qlqetrtmaeanlh-rqgeelr--klsqqlqmvhggqlqmfqgspnqlfsgvqlsg 584

OY 645 LTTSDQNEPPLASSLSDTKDKDSTGSLPGS-GSTHGTSLEKHKILRLQLQDSSPYDL 703
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 585 nsnlqqlaplmgqgvprlnqigsqmmthlygtl-----qlmldqqlqstsltsqsg 637

OY 704 AKLTAATGKDLQSESSS--TAPGSEVTIKOEPVSPKKENALLRYLKDXTKDIGLPE 761
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 638 nvlsgheqstlpsqgtqlcaplyntlmvlsqpaagsm-----vqlps 680

OY 762 ITPKLERIDSKTDPASNTKLIAMKTEKEEMSFEPGQD 798
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 681 smpq-----nstgsaaavtlf-----lqdrqirfsgqgq 708

```

```

RESULT 20
AA132215
ID AA132215 standard; Protein: 846 AA.
XX
AC AA132215;
XX
DT 15-FEB-2000 (first entry)
XX
DE Human CLOCK protein.
XX
KW CLOCK; human; transcription factor; circadian rhythm;
KW jet lag; sleep disorder; depression; seasonal affective disorder;
KW fertility; therapy.
XX
OS Homo sapiens.

```

```

XX
PN WO9957137-A1.
XX
PD 11-NOV-1999.
XX
PF 06-MAY-1999; 99WO-US10072.
XX
PR 07-MAY-1998; 98US-0084610.
XX
PA (HARD ) HARVARD COLLEGE.
XX
PI Weitz CJ, Gekakis N, Staknis D;
DR WPI: 2000-052938/04.
XX
DR N-PSDB: AA234629.
XX
PT Novel heterodimeric composition for identifying modulators used in
PS diagnosing and treating circadian clock disruption disorders -
XX Disclosure; Fig 26; 96pp; English.
XX
CC This sequence represents human CLOCK protein. CLOCK activates the
CC transcription of the perl gene when present in combination with
CC BMAL1 protein (see AA132209). The invention, based on the discovery
CC of the transcriptional mechanism regulating genes responsible for
CC the establishment and/or maintenance of the circadian clock, is
CC useful for the assay of novel drugs aimed at restoration of a
CC normal circadian cycle, the drugs being modulators of BMAL1-CLOCK-
CC mediated transcription of E-box-linked genes. The drugs are used
CC to treat conditions such as jet lag, sleep disorders, depression
CC (seasonal affective disorder) and infertility. The invention also
CC provides BMAL1 and CLOCK proteins with which to stimulate the
CC transcription of an E-box-linked gene which regulates the circadian
CC clock.
XX
SQ Sequence 846 AA;

```

```

Query Match 4.28; Score 317; DB 21; Length 846;
Best Local Similarity 21.58; Pred. No. 7.1e-12;
Matches 176; Conservative 129; Mismatches 316; Indels 196; Gaps 39;

```

```

OY 31 KRNETKRRNEOEKNYIEELAEILFANFNIDNFNKPKCALIKETVQIROIKOEKAA 90
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 39 rnksekkrtrdgnvnlkelsgmipgnar-----kmdkstvlqtsidflr--khkelta 89

OY 91 AANIDEVOKSDVSTGOGVIDKDALGPMLEALDGEFFVNVLEGNVFESENVTOYLRYN 150
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 90 qsdasel-rqdwkpc---flsneeftqlmlealdgfflaimgdtsilyseesvtallehl 145

OY 151 QEELMNKSVYSILHVGDTHEFVKNLLPKSIYNGSGMSEPPR-RNSHTFNCML----- 203
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 146 psdlvdgsifnflpegehsevyk-llstchllsdsllpeylyksknqlfecchmlrgtldp 204

OY 204 -----VKPLPDESEEGHND-----QEAHQ-KYETMOCFA-----VSOPKSIK 239
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 205 kepstleyvklfgnfxkslnsvssahngfegltqrthrpseydrctvatvrlapqfllk 264

OY 240 EBGEDLOSCLICVARRVPMKERPVLPSSSEFTTRDLOGKITSLDTSTMRAMKRGWEDL 299
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 265 e-----mctv-----eep-----neeftshlslewkflfld---hrappllylplf 302

OY 300 VRRCIOKFAHQEGESVSYAKRHHHEVLROGLAFSOIYRFLSDGTIVAQTKSKLIRSQ 359
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 303 evlgtsgdydyhvddlenlakchen-lmqyqkscyyrfltkgqgwawlqchyllyhq 361

OY 360 TTNEPOLVLSLHMLHREONVCMNPDLTGOTWKGKPLNPISNSPAHQALCSGNPGODMTL 419
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 362 wnspefivchtvsvsyaevraerrelg-----ieeslpetaadksqdsadnri 412

OY 420 SSNINFPINGPKQMGMPGRFGSGGNHNVSGMOATTPOGSNVALKMN--PSOSSPGMN 478
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 413 nt-----vsjkealerfdhs-----pfpasasrsrkshtavsdpst 451

```



QY 1168 -----LRMPPRGLRPTGLVONOPNOLLOLQHL-----QAQ 1200  
Db 1508 spapnwgagpprqaapppg-aphbpriqgpaqvagdqhcypppqgpppppqgqgq 1566  
QY 1201 QNRQPLMAQIN-----VSNVITLREPV-PTQAPINMAOLAORQELINQ 1245  
Db 1567 qqqqppqvagpppqgppqgppqagppqtaqsgiaappspilrppspgqgqrmppgm 1626  
QY 1246 HLROROMHQOQVOORFLMAMGQGLNMTPSAVAPSGMATSNPRLPQAN-----AQQ 1298  
Db 1627 paqgqsgqgqgvpq-----pppqgashgvp-----spglpvgpggmwkpypa 1671  
QY 1299 FPFPPNVCISQDPDGFATTPOSPLKSPMAHTQSPMAQSOA-----NPAYQAP 1350  
Db 1672 mppppsgqvgqvgqgpppgmmsqkppmpgqamqgqplqgqppshqphqhqhqp 1731  
QY 1351 SDI--NGMAGCGMGNSM 1366  
Db 1732 hqmpnqgtaqpgygpmpm 1749

## RESULT 22

AB67600  
ID AB67600 standard; Protein: 642 AA.

AC AB67600;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 29952.

KW Drosophila; developmental biology; cell signalling; insecticide;  
pharmaceutical.

OS Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

PA (PEKE ) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

DR WPI; 2001-656860/75.

DR N-PSDB; ABL11703.

PT New isolated nucleic acid detection reagent for detecting 1000 or more  
genes from Drosophila and for elucidating cell signalling and cell-cell  
interactions -

PS Disclosure; SEQ ID NO 29952; 21pp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent  
capable of detecting 1000 or more genes from Drosophila. The invention is  
useful in developmental biology and in elucidating cell signalling and  
cell-cell interactions in higher eukaryotes for the development of  
insecticides, therapeutics and pharmaceutical drugs. The invention  
discloses genomic DNA sequences (AB16176-AB16175), expressed DNA  
sequences (AB101840-AB16175) and the encoded proteins  
(AB57737-AB572072).  
CC The sequence data for this patent did not form part of the printed  
specification, but was obtained in electronic format directly from WIPO  
at ftp.wipo.int/pub/published\_pct\_sequences.

CC Sequence 642 AA;

Query Match 4.1%; Score 311.5; DB 22; Length 642;  
Best Local Similarity 23.0%; Pred. No. 1,1e-11;  
Matches 160; Conservative 86; Mismatches 273; Indels 177; Gaps 34;

QY 7 NTSDEPRAETRRKCCPCPOLGSPKRNTEKRNREQENKYEELAFIFANENDIDNENFK 66  
Db 5 nldkerfasre-nhc-----elerrrnkmtayltaisdmv-----ptcalark 49  
QY 67 PDKXILKEVYKQIQIIEQEKAAANIDVQKSVSTG---QGVIDKDALGPMMLAL 123  
Db 50 pdklilrmavahmkalr-----gtgnlssdqlykpsflldqelkhllleaa 96  
QY 124 DGEFFVNLK-GNVYFSENVTOYLRVNOEELMNKSVYSLHVGHTFEVKNLKRSTVN 182  
Db 97 dgflfvscdsgvlyvdsvtvplnyqsdqytslyehlpdpdrekiqelscqsqn 156  
QY 183 GGS---WSGEPFRNRSH-----FNCRLVLP-PSDEEGHNDQEAHQ----- 222  
Db 157 agrlidlksqtvkqeghsmslmqarrgflcrmrvgnvpsmvsghlrlkqnslg 216  
QY 223 -----KRYEMQC--FANVQPKS-----IKEGEDIQS--CLICVAR-----RVPKMER 261  
Db 217 psrdgtnyavvchtcglylknwptldmfpmhmerdvdmshoclaigrlyvstaadm 276  
QY 262 PVLPSSESEFTTRQDLOGKITSLDTSTMRAMKPGMEDLVRCIOKFHQHEGESVYAKR 321  
Db 277 sgsmngsefltrhamdqftrvdqrvlnllyltptellgkiclydfnp-----edslnke 332  
QY 322 HHHEVLRQ-GLAFSQIYRFLSDGTLVAQTKRSKLRSQTNEPOLVLSLHLHREQNV 380  
Db 333 sfqgylkqkgmfsllyaraknseywlrtdqayflnpyldeveyl-----vc 381  
QY 381 VMNPDLTCQTM-GKPLNPISSNSP-----AHQALCSGNPQDMTLSSNINPNGP 430  
Db 382 t---nssgklmhgablidaaahltpeqvqgqgqghvyvgaapydva----- 426  
QY 431 KEOMGMPMRFGSG-GNMNVSGMOATTPQGSNNVALKNKNSPQSSPQGNPCQPTS-MLSP 488  
Db 427 rrel-tpvgsatndmgymtlnlamqaptpq-----qqgqgqgqpgsaqgttptytydt 478  
QY 489 RHRMSPGVAGSP---RIPSQFSPAGSLHSPGVGCSSTGNSHSTNSLNL----- 537  
Db 479 th--spysaagqpplaklpkqgstp-----cpv-----apnswaalrpgqgq 519  
QY 538 --QALSECHGVSLGSSLASP-----DLKMGN-----LDNSPVMNMPPLSKNGS 579  
Db 520 qqqpvttegyyqqtsparspsqptytqlsaqngnrqgaqgaqgpp-----pppnabgm 575  
QY 580 LDKKDCFLYGEPPSEGTGQAESSCHPGEQKETNDP 615  
Db 576 wdqgqagq-hpnhphtaphhphahpypagaqgp 610

## RESULT 23

AAW79529  
ID AAW79529 standard; Protein: 855 AA.

AC AAW79529;

DT 02-FEB-1999 (first entry)

DE Mouse CLOCK protein.

DE CLOCK; circadian rhythm; mouse; jet-lag; sleep-wake disorder;  
seasonal affective disorder; cancer; transcription factor;

DE diagnosis; therapy.

OS Mus sp.

XX Key Location/Qualifiers

FT Protein 2..855 /note="Claimed fragment"

FT Protein 10..855



transcription of the *per1* gene when present in combination with BMAL1 protein (see Aki32209). The invention, based on the discovery of the transcriptional mechanism regulating genes responsible for the establishment and/or maintenance of the circadian clock, is useful for the assay of novel drugs aimed at restoration of a normal circadian cycle, the drugs being modulators of BMAL1-CLOCK-mediated transcription of E-box-linked genes. The drugs are used to treat conditions such as jet lag, sleep disorders, depression (seasonal affective disorder) and infertility. The invention also provides BMAL1 and CLOCK proteins with which to stimulate the transcription of an E-box-linked gene which regulates the circadian clock.

**SQ Sequence 855 AA;**

Query Match	4.18;	Score 309.5;	DB 21;	Length 855;
Best Local Similarity	21.78;	Pred. No. 2.2e-11;		
Matches 150;	Conservative 107;	Mismatches 254;	Indels 181;	Gaps 27;

```

OY      31 KRREKNNREDEKNYIEELMELIFANRNDIDNENFRKDKCALIKETVAKTORIKREQKKA 90
Db      39 rnksekrttdfnvllkelsgmipgnar-----kmdkstvlqskidflr--khxetta 89
OY      91 AANIDEVOKSDVSSTGGGVIDKDKALGPMALDGEFFVNVNLEGNVVFSENNOTYOLRYN 150
Db      90 qdsdel-rqdwkpt---flsneefctlmleadqfllamtqgsilysesvstsllehl 145
OY      151 QEELAMKSVSILHVGHTPEVKMLRKSLYNGSGWSGGEPPR-RNSHTFNCRL----- 203
Db      146 psdlvdsilfnfipegehseyuk-llsthlledsdlpreyiksxnqjefctchmlrgrtdp 204
OY      204 -----VRPLPDEEGHNN-----OEAHQ-KYETKQCPA-----VSOPSRK 239
Db      205 kepstyeyvrlfngfkstlstshtngfegrlqtrhrpyedvctvalvrlatpbfik 264
OY      240 EEEGDLSCLCVARRVPMERPVLPSESEFTTRDLOGKITSLDITSTMRAMPGEWDL 299
Db      265 e-----mctv-----eep-----neefsthslewklfld--hrappilgylpf 302
OY      300 VRRCIOKFHAQHEGESVYAKRHHEHVLROGLASQIYRKSLSGTLVAAQOTSKILRSQ 359
Db      303 evlgtsgdydyhhvddlenlakchen-lmqy9k9k9scyryfltky9q9w1qtltyllynq 361
OY      360 TTNEPOLVILSHMLHREQONVCVNMNDLTGTGMKPLNPISSNSBPAHQALCSGPNQDWTL 419
Db      362 wnspefivcthtvsvyaevraerrelg-----leeslpetaadksqsgsdnrl 412
OY      420 SSNNINPINCXEDGMKPMGRFSGSGGMNYSGQATTPGGSNTALKMNS-PSQSSFGMN 478
Db      413 nt-----valkealerldhs-----plpaasrstrsksthtvsdpst 451
OY      479 PGQ-PTSMLSP-----RHRMSGVAGSPRIEPSQFSPAGSLHSEVGCSSR 523
Db      452 pltkiprtclstpprqhlpahekmtgrstsfssqgsinsqsvgspltkpmsqaanlpl-pq 509
OY      524 GNSHSYTNSSNALQALSE-----GHEGVSUAGSSL 552
Db      510 gmsqfqtgsaqigamqhlkdqlqgrtrtmleanlhrqgeelk1qeqdlqmwngq1lqmtlqg 569
OY      553 ASDPLKMGKGNLONS-----PVNNKMP-PLSKSGSLDSKDCFCGLYGEPEEG----- 595
Db      570 snpplnfgsvqslssgnsnlqglprvmngqgvvpanqygs-----ghlstqgmhmq 619
OY      596 -----TTGOAESCHAPGEOKETENDNLPFA 620
Db      620 qqt1qststq9q9qsvm9shgn9qstslpsq9s 651

```

RESULT	25
ABB68896	
ID	ABB68896 standard; Protein; 2439 AA.
XX	

AC	AB66896;
XX	
DT	26-MAR-2002 (first entry)
XX	
DE	Drosophila melanogaster polypeptide SEQ ID NO 33480.
XX	
KW	Drosophila: developmental biology; cell signalling; insecticide; pharmaceutical.
KW	
OS	Drosophila melanogaster.
XX	
PN	WO200171042-A2.
XX	
PD	27-SEP-2001.
XX	
PF	23-MAR-2001; 2001WO-US09231.
XX	
PR	23-MAR-2000; 2000US-191637P.
PR	11-JUL-2000; 2000US-0614150.

W0200171042-A2.

27-SEP-2001.

23-MAR-2001; 2001WO-US09231.

23-MAR-2000; 2000US-191637P.

H  
H  
C  
C  
E  
R  
C  
C  
C  
C  
  
C  
C  
H  
W  
H  
U  
C

11-11-11

1  
 2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525

N-PSDB; ABL12999.

New isolated nucleic acid detection reagent for detecting 1000 or more

interactions -

Disclosure; SEQ ID NO 33480; 21pp + Sequence Listing; English.

The invention relates to an isolated nucleic acid detection reagent

useful in developmental biology and in elucidating cell signalling and

insecticides, therapeutics and pharmaceutical drugs. The invention

sequences (ABL01840-ABL16175) and the encoded proteins

The sequence data for this patent did not form part of the printed

at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://ftp.wipo.int/pub/published_pct_sequences).

Sequence 2439 AA;

Query Match	4.0%;	Score 307;	DB 22;	Length 2439;
Best Local Similarity	21.0%;	Pred. No. 1.7e-10;		
Matches 280;	Conservative 152;	Mismatches 486;	Indels 414;	Gaps 63

```
Oy 280 ITSLEDTSTMRAMKPGMEDLVR-----RCIQKFHAQHEGSEV-----S 31
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 5 lqglepslvrlvlslpwwdaqrtrqlandgnpelralnigsvqfegdsvinlkvvggdik 64
```

QY 318 YAKRHHHEVL-----RQGLAFSQIYRFSLSDDTLVAAQTCKSLIRSQTTNEPQVLI 368

Db 65 itkdnvnetleaaglsnksylatgfdgpbstksaasafmqgqqrllqpnd--ai 121

QY 369 SLHM-----LH-----REQNCV-----MNPDLTGQTMGKPLNPIS 400

Db 122 slqrraaaglg1qkmp1ht1tqqqqqqqqvavtpavfksptvcpmegkv--p111 175

QY 401 NSPAHQALCSGNPGQDMTLLSNINFPINGPKEQMGMPMGRFGGSG-----GMNH 4495

Db 180 psp-----satrdfpfesmrgarvlgreaaglgpplppppppnvtlkv 223

QY 450 VSGM<sup>Q</sup>ATTP<sup>Q</sup>GSNYALKMNSP<sup>Q</sup>SSPGMNP<sup>Q</sup>GPT<sup>Q</sup>SM<sup>L</sup>SP<sup>R</sup>HRMSP----- 494

Db 224 lktqgqapq--ngeltptptsap-tppsksqfiqpppppyglgaatassvspia 280

QY 495 --GVAGSPRI--PPSQFS-PAGSLHPVGVCSSTGNSHSYTNSSL--NALQALSEGHGVS 547





```

OY 189 EPPRRNSHTFNCRLMLKPELPDSEEGHNDQEAHQKYTEMOCFAVSGPKSIREEG----- 242
DB 157 -----kxskdmsterdfmfmkctvtgrgtvnlkstwtvyl 193
OY 243 -----EDLOSCLICVARRVPKKEPVLPR-SSESFTTRDLOGLK 279
DB 194 hctgqvkvynncpnhslcgykeplscilimceplqphshmdlpldsktflsrhmdmk 253
OY 280 ITSLSDTSTMRAMKRCW--EDLVRCIOKFNHQHGESVSYSKRRHHHEVLRGLAFSQT 337
DB 254 ftycd-----rtteligyhpheelgrsayefynaldsemt--kshqnlctkygvvsqy 307
OY 338 FESLSDTGLVAQAQTSKILRSQTNELPOLVLSLH--MLHREONVCMVMPDLTGGTMGRP- 394
DB 308 rmlakhggyvyletqgtvlylprnlqpgcimevnyviseiekndvfmtdt-aelflkph 366
OY 395 ---LNPISNSPAHQALCSGNPGDMLSSNINFPINGPK-----QMGMPMG----- 439
DB 367 lmannsifds-----gky-avsekenflftklkeepeelaqlaprgdaalsl 414
OY 440 RFGSGGNHNSGMQATTPOGSNVALKMNPS--QSSPGMNPQPTSMLSPRHRMSPGVAG 498
DB 415 dfgnqfneessaygkailprsqwateltrshstgseagslpafv-----pqaaa 464
OY 499 SPRIPPOFSPAGSLHSPVGVCSSTGNSHSYTNSLSNLQALSEGHSGLASPLDK 558
DB 465 pgrtstpsatssss-----cstprnspeydtysldndlk-----levleklfamdte 510
OY 559 MGNLONSFVNMMNPPLSLKMSGLSDKDCFGLVGPESEGTGQAESSCHGEOKETNDPILP 618
DB 511 akdgcstqtdfnejdleclapy-----ipmdgedfqlspc-peerllaenpss 559
OY 619 PAVSSERADGOSRLHDSGQR-----KLQLLTTSDDMEPPLASS 660
DB 560 pq-----hcfamtnlflplavaphspflldkftqgleskktepethrpsmsl 607
OY 661 LSDTNKDSGSLP--SGSGTHGTSLKKEHKILHRLDSSSPVLAKLRTAATGKDLSDQ 717
DB 608 ffdag--skaslppccgqastplsmggrsn-----tqwpdpplhifpklway-dqrt 659
OY 718 ESSSTAPGSEVTIKOEPVSPKKENALLRYLLDKDTRDGL--PEI-TPKLERLDSKTD 774
DB 660 eflgaap-----lgyppvspbhvstfkr-----sakfgargpddvlspmal- 702
OY 775 PASNTKILAMTEKREMSFER--ODRGSELNDLE-ELLDLONSQPLPFPDRPGAP 830
DB 703 --snklklrqlgeyeeqdfqldsgdppgystshlmwkrmklrlygscp-lmpd----- 753
OY 831 AGSVDKQAIINDLMQLTAENSPVPVGAQKTALRISQTFNNPRPGOLPRLNQNLPD 890
DB 754 -----xpisanvndkftg-----nprug-1gh-plthldlp 783
OY 891 ITLOSPTGAP-----FPIRNSSPYSVLPQGMNGOQMGNGOQMGNGOQMGNGOQ 943
DB 784 ---qppsaipsgenskrifpqcayatyq-----dyslssakhsqgm 822
OY 944 ASRPMPGEMAPQSSANRVCAATTSAMNRPVOC 978
DB 823 asrlipgsfesyllpeltrydc-----evnvpvlg 852

```

RESULT 27

AA79161

AA79161 standard; Protein; 870 AA.

AA79161;

05-JUN-2000 (first entry)

Human endothelial PAS domain protein-1.

Endothelial PAS domain protein-1; human; EPAS1; angiogenesis;

antiartherosclerotic; antitumor; atherosclerosis; tumour;  
vascular disease; vulnerability; cardiatic; vasculogenic;  
cerebroprotective; gene therapy; transcription factor.

Homo sapiens.

Key Location/Qualifiers  
FH 14..67  
FT Domain /note="N-terminal bHLH domain"  
FT Domain /note="transactivating domain"  
FT Domain /note="transactivating domain"

WO200009657-A2.

24-FEB-2000.

13-AUG-1999; 99WO-US18539.

14-AUG-1998; 98US-0096515.

(HARD) HARVARD COLLEGE.

Lee M, Maemura K, Hiesh C;  
WPI; 2000-205996/18.  
N-PSDB; AA294053.

Modulation of angiogenesis in mammals, useful for treating e.g. atherosclerosis, tumors, wounds, vascular disease, hypoxic tissue damage, ischemia, balloon angioplasty, frostbite, gangrene or poor circulation

Claim 11; Page 13-14; 57pp; English.

The present sequence is that of human endothelial PAS domain protein-1 (EPAS1), a member of the transcription factor family characterized by a basic helix-loop-helix (bHLH) domain and a Per-ARF-Arnt-Sim (PAS) domain. The invention is based on the discovery that EPAS1 binds to cis-acting regulatory sequences associated with genes encoding angiogenic factors such as vascular endothelial cell growth factor (VEGF) and VEGF receptors such as KDR/Flk-1 and Flt-1, thereby transactivating the promoters of such genes. A claimed method of inhibiting angiogenesis in a mammal comprises administering to the mammal a compound which inhibits binding of EPAS1 to the cis-acting transcription regulatory DNA of an angiogenic factor (see AA294051). The compound may be an EPAS1 polypeptide lacking a transactivation domain (see AA79160) or a nucleic acid encoding such a polypeptide, an antisense nucleic acid complementary to mRNA encoding EPAS1, or an EPAS1-specific antibody. The compound is preferably administered to an atherosclerotic lesion or to a tumor site. Angiogenesis is also inhibited using a compound, such as an EPAS1 dominant negative mutant, which inhibits binding of EPAS1 to the EPAS1-binding element, ARNT4 (see AA79162). Angiogenesis can be promoted by administering EPAS1 DNA to increase expression of VEGF or VEGF receptor in endothelial cells of a patient suffering from peripheral vascular disease, cerebral vascular disease, hypoxic tissue damage (e.g. hypoxic damage to heart tissue), or coronary vascular disease as well as to treat patients who have, or have had, transient ischemic attacks, vascular graft surgery, balloon angioplasty, frostbite, gangrene, or poor circulation.

Sequence 870 AA:

Query Match 3.8%; Score 292; DB 21; Length 870;

Best Local Similarity 20.6%; Pred. No. 3.2e-10;

Matches 217; Conservative 138; Mismatches 406; Indels 294; Gaps 46;

10 DPSRAETRRKRECDQLOPSFRNTEKRNROENKYIEELA-ELIFANFNDIDNFPKPD 68

6 eksrsserirke-----ksrdarcrirkelevfeyelhelplph-----svshld 52

```

QY 69 KCAILKEVYKQIQIKOEKQEAANAANIDEVOKSDVSTGCGIDKDALGPMMLBALDGFEE 128
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 53 kasimleisflrtkhllsvcseneseead-----qgmndilylkallegfia 100
QY 129 VVVLBNVAVFVSNVNOYILRYNOELMKNKSVYLHVGHDFEVKNLPLPSYNGSGMSG 188
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 101 vvvqgdgmflfiseniskfmgltqveltqhsifdfctpcdeletrnl--slkngsfq- 156
QY 189 EPPRRSHTEFNCMLVPLPDESEEGHDNOEAHQKYEYTMQCFVVSQPKSIKEG----- 242
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 157 -----kkskdmsterdfmfmktctntnrgtrvnlksatkv 193
QY 243 -----EDLOSCLICVARRVPMKERYLP--SSSEFTTRDLOLQK 279
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 194 hctgvykvynncpphnlscykeprllsccllmccepqhpsmdiprldsktflsrhmdmk 253
QY 280 ITSLSDTSTRAAKPGM--EDLYRCLQKFNHAGESESVYARHHHEVLRQGLASQIT 337
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 254 ftycdp---rtel1qyhpeellgrsaeyelghldeennr--ksnqnclctkgvvsqgy 307
QY 338 RFSLSGDTVAQAOTKSKLIRSOYTNNEPOLYISLH--MLHREQNVCVWNPDLTGOTMGKP- 394
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 308 rmlaknggywvletgtvlynpnlqpqclmcvnyvlselektndvtfsmqt-eslflkph 366
QY 395 ---LNPISNSPAHQALCSGNPGQDMLSSNINFPINGPRE-----QMGMPMG----- 439
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 367 lmmamslfdds-----gkg-avseksnflftklkepeelaqlaptpgdalisl 414
QY 440 RFGSGGMNHSVGMQATTPGGSNVYALKMNSPS--QSSPGMNPQPTSMLSRHHMSPGVAG 498
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 415 dfngnqfreesaaygkallpssgpwalelrshscqseagslpaftv-----pqaa 464
QY 499 SPRIPOSEFSPAGLSLSPVOCSSSTGNSHSYTNSSLNALQALSEHGVSLSGSSLASPDLK 558
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 465 psgtspatssss-----cstpnspedytslndnlk-----levleklfmdtre 510
QY 559 MGLQNSPVMNMPPLSKMKSLSKDCFGLYGPESEGTGQAESSCHPGQKKTNDPMLP 618
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 511 akqcsctqdfideldletlapy-----lpmqgedfqlsplic-peorlllaenpqst 559
QY 619 PAVSSRRADQGRRLHDSKQGT-----KLQLLTKKSDQMEPSPPLASS 660
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 560 pq-----hcfamtnlfgplavaphspflldkfqqleskktepehrpmsl 607
QY 661 LSDTNKDSGTSLP---GSGSTHGTSLKEKHKLHRLLDSSSPVDLAKLTAEATGKDLQ 717
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 608 ffdag--skaslppccgqgstplsmggrsn-----tqwpdpplhfgpkwavy-dqtr 659
QY 718 ESSSTAPGSEVITKQEPVSPKKENALLRYLDDKDTKDIGL--PEI-TPKLERLDSKTD 774
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 660 elfgaap-----lppvpapbhvstfktl-----sakfgargpavlspanaval----- 702
QY 775 PASNTKLIAWKTEKEMSFEP---GPOGSELBNLE-ELIDDLQNSQLPOLPFDTPRGAP 830
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 703 --snklklkrqleyeeqafqdlsgdpppgstshlmwktmknltrgscp-lmpd----- 753
QY 831 AGSVNDKQAIINDMLQTLAENSPTVPVGAQKTALRLISOSTFENRPRPOLGRLHPLNQLPLD 890
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 754 -----kplsanvndkftcy-----nprmg-1gh- plrlhplp 783
QY 891 ITLQSTGAGP-----FPPINSSPYSIVTPOPGMNGMNGMNGMNGMNGMNGMNGMNGMNG 943
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 784 ---qpssaispgenskstrlppqcyatcyq-----dyslssahkvsvsm 822
QY 944 ASRPVSGEMAPQSSAVRYTCATTSAMNRPVQ 978
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 823 asrlllgpstsyl1lpeltydc-----evnvpvlg 852

```

RESULT 28  
 ABB62819  
 ID ABB62819 standard: Protein; 1963 AA.  
 XX

```

AC ABB62819;
XX
XX 26-MAR-2002 (first entry)
DT
XX Drosophila melanogaster polypeptide SEQ ID NO 15249.
DE
XX Drosophila melanogaster polypeptide SEQ ID NO 15249.
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX Drosophila melanogaster.
OS
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
PF
XX 23-MAR-2000; 2000US-191637P.
PR
XX 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE ) PE CORP NY.
PA
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
PI
XX WPI; 2001-656860/75.
XX
XX N-PSDB; ABL06922.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT interactions -
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX Disclosure; SEQ ID NO 15249; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
XX sequences (AB161840-AB16175) and the encoded proteins
XX (AB16737-AB172072).
XX
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 1963 AA;
SQ

```

Query Match 3.8%; Score 289.5; DB 22; Length 1963;  
 Best Local Similarity 20.8%; Pred. No. 1.7e-09;  
 Matches 265; Conservative 146; Mismatches 510; Indels 353; Gaps 55;

```

QY 399 SSNSPQHQAALCSGNPGQDMLSSNINFPINGPREQGMMPGRGSGGMNHSVGMQATTP 458
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 155 nmnrprrrrgqalrrnrvlranmsnntppkkedlrep-----qatata 199
QY 459 OGSNVYALKMNSPQSSPGMNPQPTSM-----SPRHMSPGVAGSPRIPOSP 507
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 200 taataas-etasarpsavsksensltdeddascdsdslltkrdes-----psrm 250
QY 508 SPAGSLHSFVQCSSSTGNSHSYTNSSLNALQALSEHGVSLSGSSLASPDLK--MGNLON 564
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 251 rtrnkqnnmsstsgnntagngnatslsgstgggaagns--sskdqsanavangkr 309
QY 565 SPVNMNMPPLSKKGSLSKDCFGLYGPESEGTGQAESSCH-----PEQKKTNDP 615
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 310 pkrgetpvdvsgavsd-----pkrptkavassanrxkqgkqdtppkkktege 360
QY 616 NLPPAVSSERADQGRRLH-DSKGQTKLLOLLTTKSDQMPSPPLASSLSDTNKDSGTSLG 674
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 361 snepsaheenalkekrrkpdspve-----smnsdsrpsvlddgsnttttttaegqs 413
QY 675 SGSTHGT-SLKEKHKLHRLLDSSSPVDLAKLTAEATGKD-----LSQESSSTAG 725

```



```

Db      214 rpehprlgkvcfclatvrlatpqlke-----mcva-----depl-----eeftar 254
QY      274 ODLOGKITSLDHSTWMAANKPGM---EDLVRRCIQKFNHQHGESVYAKRHHHEVLRQG 330
Db      255 hslwffild---hnppllgylpfevlgtsgynyhl-----dellarchqhmqlg 307
QY      331 LAFSQIYRFSLSDGTLVAAQTCKSLKILRSQTNPEQLVLSH-----MLHREQNVCM 382
Db      308 kskscyrfltcgqgwltwqthlyltlyhgmkskrelcthvsvadydrvertgelale 367
QY      383 NPDLTGQTMGKPLNPISNSPAHQALCGNPGQDMTLSSNINPINCPEQMGPMGRFG 442
Db      368 dp-----pleam----- 374
QY      443 GSGGMHVSGMQATTPOGSNYALKMNSPQSSPGMNGOPTSMLSPRHMSPGVAGSPRI 502
Db      375 -----hpsavk-----ekdssleppqpfal-----dmgaaglpss 405
QY      503 PPSQFSPAGSLSHSPGVCSSTGNSHSYTNLSNALQALSEGHGVSLSGLSASPDLKMGNL 562
Db      406 p-----spasars-----shkshtams----- 424
QY      563 QNSPVNMNPPPLSKMGLSDSKDCEFLYGER-SBGTGQAESSCHEQEKEETNDPNLPAY 621
Db      425 -----eptstptklmaenstlajpratlpgelprvqg 456
QY      622 SSERADGOSRLDSSKG---QTKLQLLTTSKSDQMEPSPLAS-----SLSDNKKDSTGSL 672
Db      457 lsgaaampalhsaaadltkqlllqslpqtqglsppavvtqfsaqlmfqtkld----- 511
QY      673 PGSGTHTGTLSEKHKILHRLLDSSSPVDLAKLTAEATGKDLSDSSSTAPGSEVTIKO 732
Db      512 -----qlgeqrtrllq-----anlrwgq 528
QY      733 EYVSPKKENALLRYLLDDDKDIDGLPEITRKLERLDSDTDPASNTKLIAMTEKEEMS 792
Db      529 eelhkiqueqlclvq-----dsnvqmf-lqgavvsis 558
QY      793 FEPGDQPGSELNLEETLDDIQLNSQLPOLFPTRPCAPAGSVDKQATINDLMQLTAENSP 852
Db      559 fssatqppaaq-----qqlqg-----rpaaps-----qpqlvvn----- 586
QY      853 VTPVGAQKTALRISQS-----TFNNPRPGQLGRLLPNQNLPL-DITLOSPTGAGPF 902
Db      587 -lplgqgltstqvtqthqllresnvlsaqgkpmrseqllpasgrslslpsqfscstasvl 645
QY      903 PPIRNSPPTSVIPQPGMNGOGMIGNGLGNS-STGMIGNSASRPTMSEGAAPQSSAV 961
Db      646 pqlsltltlaptpq-----dsagcqpssdfghdrqlrlllsqplqpmmpqscdarpsav 700
QY      962 RVT-----CAATTSAMNRPVOGGMIRNPASIPRPSQPGOROTLQS 1004
Db      701 strtgrykvkagqymfmpspdshtpsastprv---llmgqavlhpsfprsrpqlg----- 753
QY      1005 QVANNIGPSELMMNGGPQYSSQQAAPPNGTAP-----WPESILPIDQASFASONKOPFS 1058
Db      754 -----paqagqgpppylqaptalhsqpsdl-----lftsfqppqlgy 793
QY      1059 SPDDLCPHPAAFS 1072
Db      794 aatqslcpqpprps 807

```

```

RESULT 30
ABB62231
ID ABB62231 standard; Protein; 2441 AA.
XX
AC ABB62231;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 13485.

```

```

XX      KW Drosophila; developmental biology; cell signalling; insecticide;
KW      OS Drosophila melanogaster.
XX      PN WO200171042-A2.
XX      PD 27-SEP-2001.
XX      PF 23-MAR-2001; 2001WO-US09231.
XX      PR 23-MAR-2000; 2000US-191637P.
XX      PR 11-JUL-2000; 2000US-0614150.
XX      PA (PEKE ) PE CORP NY.
XX      PI Venter JC, Adams M, Li PWD, Myers EW;
XX      DR WPI, 2001-656860/75.
XX      DR N-PSDB; ABL06334.
XX      PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT      genes from Drosophila and for elucidating cell signalling and cell-cell
PT      interactions.
PS      Disclosure: SEQ ID NO 13485; 21pp + Sequence Listing; English.
XX      CC The invention relates to an isolated nucleic acid detection reagent
CC      capable of detecting 1000 or more genes from Drosophila. The invention is
CC      useful in developmental biology and in elucidating cell signalling and
CC      cell-cell interactions in higher eukaryotes for the development of
CC      insecticides, therapeutics and pharmaceutical drugs. The invention
CC      discloses genomic DNA sequences (ABLI6176-ABLI30511), expressed DNA
CC      sequences (ABLI01840-ABLI6175) and the encoded proteins
CC      (ABB57737-ABB72072).
CC      The sequence data for this patent did not form part of the printed
CC      specification, but was obtained in electronic format directly from WIPO
CC      at ftp.wipo.int/pub/published_pct_sequences.
XX      CC
XX      SQ Sequence 2441 AA;

```

```

Query Match 3.8%; Score 287.5; DB 22; Length 2441;
Best Local Similarity 19.2%; Pred. No. 3.2e-09;
Matches 296; Conservative 195; Mismatches 595; Indels 453; Gaps 61;

```

```

QY      6 EMTSDPSR---AETPKRKRCPDQLGSPKRTKRNREOKNYIELLAFANPDID 61
Db      1136 ertresrkaaaarrrerk-----kkmeekkeekrrqgngpggdmdqgdddasdkd 1190
QY      62 NFNFKPDKCALIKETVKQIRQIKEDKKAANIDEVQKSDVSTGCGVIDKDAL--GPMK 119
Db      1191 ddsdkd-----edeaapaaree-----gdsjldqscsagdk 1226
QY      120 LEALDGEFFVNVLEGNVVFSENVTOYLRYNOBELMKNKSVYSLIHVGHTFEYKNILPKS 179
Db      1227 garfgsgsaagaanaavstn--sqgkknkqaknkylisv---eptqpyv--ltsns 1278
QY      180 IYNGGSMSEPP-----RRSHTFNCRMLVYKPLPDSSEEGHDNOEAHQKETMOCFAVNSP 235
Db      1279 vlkyvcaakhhpavevkvqpapatqgaaplkrqldavkkee-----palkkkekns---sss 1330
QY      236 KSIKEGEDLQSLICIVARVPMKERP-----VLPSSEFTTRQDLQKITSIDTSTMRAL 291
Db      1331 ssakreenlapekva-----lpkqgsssssklgsesasnlstatltsaantlrkey 1386
QY      292 MKP-----GMEDLVRRCIQK-----FHAQHGESVSYA 319
Db      1387 akpasqtasatltnpakrtvvgkwkksaagqltavgaagaprlpvatatsatsvqh 1446
QY      320 KRHHHEVLROGLAFQIYRFSLSDGTLVAAQTCKSLKILRSQTNPEQLVLSLHMLHREQNV 379

```

```

Db 1447 phhh-----lans-----ssnssslttsttaassv----- 1473
QY 380 CVMNPDLTGTMGRKPLNFISNSPAHQALCSGNPGQDWTLSNINFINPGCKEQGMGMDG 439
Db 1474 -----pencckkvqpvnaiss----- 1489
QY 440 RFGSGGNNHVSQMATTPOGSSNVALKMNPSQSSPCGNPGOPTSMLSPRRHMSPGVAGS 499
Db 1490 rviqrgsgn-inairatl-gahlevekqknqserciltkyltdatkgahmliallkd 1546
QY 500 PRIPPSQSPAGSLSPVGVCSSTGNSHSYTNSSLNALQALSEGIVGLSGSLASPLDKM 559
Db 1547 pdvdlqmlpr-----insslik-qaassgg-----astpmav 1576
QY 560 GNLMN-SPVNNNPPLSKMGSLDSKDCFGLCGEPESEGT--GOAESCHPEKOKETNP 616
Db 1577 gtwcnrtcaagynaytfssaaatstss-----sssaastltpgaasyshahkqhqqpgsvk 1632
QY 617 LPPAVSSE--RADQS--RLHDSKGOTKLLQLLTTRKSDMPEPSPLASSLSDTNKDSGSLP 673
Db 1633 gpgsgstsvksngsstkvsasgs-----gsrsgagssyla-----qqqprss 1678
QY 674 GSGSTHGTSLKEKKHKLRLHQLDSSSPVDLAKLAETAGKLSOBSSSTAGSEVTTKOE 733
Db 1679 gggssngv-llskks-----essskslpaakskstlqk-----sstlvspgaqnfaakaa 1725
QY 734 PV---SPKKRENALLRYLLDKDKDGLPELTPKLERLDSKTDP----- 776
Db 1726 aigsgspkkaegatsavvtcaggrsgsvapfgrgkpvaggggpaataasnavqlgsvs 1785
QY 777 -----SNTKLIA----- 783
Db 1786 gsnllagpigtfnvadvaavnaaagaataasnvkrlapiapskkrvgspitqvqgh 1845
QY 784 -MKTEKEMSEPEPDQSGSELDNLEHLLDQLNSQLPOLPFD--TRECAPPGS-----YDK 836
Db 1846 qtcqqqqqqqlpqpapvypqpq--qbp1qgqqqqqapqpqbpqbpqbpqtsqnlvin 1903
QY 837 QAILNDLMOLTAENSPTVPVGAOKTA--LRISQSTFNNRPOLGRLLPQNQLPIDITLGS 895
Db 1904 tnllndlmaasaanttsdsfaqlaaklssayslfsdyqsgqgk1----- 1949
QY 896 PTGAGPPPLTINSPPYVIOPGMGNOGMIGN--QCNLGNSTGMIGNSASRPTMPSG 952
Db 1950 -----qdpq1gqgagavdg1pqad-askapgyrnrlilsspygssk 1989
QY 953 EMAPQSSAVRYTCAATISAMRPFVQGMIRNPASIPMRBSQSGQKOTLSQVYNIIGPS 1012
Db 1990 assnhsctspvgunv1qgqqqqqpqssqqa1nlltspgypgapatarspmwsaneqn--pa 2047
QY 1013 ELENNMGCPQYSQOQAP---PNOTAPWPESTILPIDQASFASQONQPPGSSPDDILCP--- 1066
Db 2048 vqgsmtgctq1getapahspgv1kp-pta1vpl-----qthvmp1isepegapptcf 2099
QY 1067 -----HA-----AESPSDEGALLDLYLALRNFGLLEIEDRALG1PELVQSOSQ-- 1110
Db 2100 gajsgnpsasgnsaagaataaasamltdcqqnlqltqlnq1mrvgsaqqqpqq1 2159
QY 1111 --ANDP--EQSSQDSNIMLEQKAVFPQOYASQAMOGSISPMQDNFRHMGCRPSYAT 1167
Db 2160 nypmcdpsst1vdannv1rl1nprv1fp-----qgn1kppqp-----pqggt 2201
QY 1168 LR--MQRPGILRPTGLVQONOPNQLRLQ-----LOHRLQAOQONQPLMNOISNVSNVLT 1219
Db 2202 qsnvfgnprqprq1garq1pggaaagrwyg1tleypsygrdm1lhengaggaagm--- 2258
QY 1220 LRQGVPTQAP1NQMOMLQORQREILNQHLQRQMHQOQOVQORTLMMGQGLNMTPSNAP 1279
Db 2259 ---gspsamspnhd1rkmprijtge-----raaswkyntfnvgps1medala-- 2305
QY 1280 SGMFATSNRPRIPOANNOQCFEPFNNGISQOPDGFGTGATTPOSLKSPRAHNTQSPMQ 1339
Db 2306 sv1pvpwaelkagppq1qgppppqsgqqqqlnw-----lkqbpqgqyraynngppq 2361

```

```

QY 1340 QSOANPAYQAPSDINGMAQGNMGNSMFSQSPPHFGQO 1378
Db 2362 qgqgqhepmmpmdyh-----nmqapnmasq 2387

```

```

RESULT 31
ABB31417
ID ABB31417 standard; Peptide; 149 AA.
AC ABB31417;
DT 01-FEB-2002 (first entry)
DE Peptide #4068 encoded by breast cell single exon nucleic acid probe.
XX Human; microarray; single exon probe; gene expression; breast;
KW disease; cancer.
OS Homo sapiens.
XX WO200157271-A2.
PN 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US00662.
PF 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
PA (MOLE-) MOLECULAR DYNAMICS INC.
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-496933/54.
XX New spatially-addressable set of single exon nucleic acid probes,
PT useful for measuring gene expression in sample derived from human
XX breast, comprises number of single exon nucleic acid probes -
XX
PS Claim 27: SEQ ID NO 14385; 327pp + sequence listing; English.
XX
CC The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human breast and BR 474 cells. The method involves contacting
CC the probes with a collection of detectably labelled nucleic acids
CC derived from mRNA of human breast, and then measuring the label
CC bound to each probe of the microarray. The probes are useful for
CC verifying the expression of regions of genomic DNA predicted to
CC encode proteins. They are useful for gene discovery, and for
CC determining predisposition and/or prognosing breast disease. Gene
CC expression analysis is useful for assessing the toxicity of chemical
CC agents on cells. The microarray of this invention presents a far greater
CC diversity of probes for measuring gene expression, with far less bias
CC than expressed sequence tag microarrays. The method is suitable for
CC rapid production of functional information from genomic sequence. The
CC present sequence is a peptide encoded by a single exon nucleic acid
CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pcl_sequences.
XX
SQ Sequence 149 AA;

```

```

Query Match 3.88; Score 287; DB 22; Length 149;
Best Local Similarity 44.28; Pred. No. 4.4e-11;
Matches 61; Conservative 31; Mismatches 42; Indels 4; Gaps 2;

```

```

Qy      653 EPSPLASSLSDTNKDGSLPGSGSTHGTSLEKHKILHRLLODSSPYDLAKLTAEATG 712
      11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
      5 esssvtspgsvsstsgvstsmhbgslqekhrilhkgnspaeavakiteatg 64
Db
Qy      713 KDLSSSSSTAPSGEVTIKQEPVSPKKKE-NALLRYLLDKDDTKDIGLPETPKLERLDS 771
      11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
      65 kd--tsstlscgdgnvqvqeqispkpkennallryllldrdpsdalskelqpqvadvn 121
Db
Qy      772 KTDPAANTKLIAMKTEKE 789
      11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
      122 kmsqctsstipsssgqkd 139
Db
RESULT  32
ABB36630
ID      ABB36630 standard; Peptide; 149 AA.
XX
AC      ABB36630;
XX
DT      04-FEB-2002 (first entry)
XX
DE      Peptide #4136 encoded by human foetal liver single exon probe.
XX
KW      Human; foetal liver; gene expression; single exon nucleic acid probe.
XX
OS      Homo sapiens.
XX
PN      WO200157277-A2.
XX
PD      09-AUG-2001.
XX
PF      30-JAN-2001; 2001WO-US00669.
XX
PR      04-FEB-2000; 2000US-0180312.
XX
PR      26-MAY-2000; 2000US-0207456.
XX
PR      30-JUN-2000; 2000US-0608408.
XX
PR      03-AUG-2000; 2000US-0632366.
XX
PR      21-SEP-2000; 2000US-0234687.
XX
PR      27-SEP-2000; 2000US-0236359.
XX
PR      04-OCT-2000; 2000GB-0024263.
XX
PA      (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI      Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR      WPI; 2001-483447/52.
XX
PT      Human genome-derived single exon nucleic acid probes useful for
XX
PT      analyzing gene expression in human fetal liver -
XX
PS      Claim 27; SEQ ID NO 29265; 639pp + sequence listing; English.
XX
CC      The invention relates to a single exon nucleic acid probe for
XX
CC      measuring human gene expression in a sample derived from human foetal
XX
CC      liver. The single exon nucleic acid probes may be used for predicting,
XX
CC      measuring and displaying gene expression in samples derived from human
XX
CC      fetal liver. The present sequence is a peptide encoded by a single exon
XX
CC      nucleic acid probe of the invention.
XX
CC      Note: The sequence data for this patent did not form part of the
XX
CC      printed specification, but was obtained in electronic format directly
XX
CC      from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ      Sequence 149 AA:

```

```

Query Match      3.8%; Score 287; DB 22; Length 149;
Best Local Similarity 44.2%; Pred. No. 4.4e-11;
Matches 61; Conservative 31; Mismatches 42; Indels 4; Gaps 2;

```

```

Qy      653 EPSPLASSLSDTNKDGSLPGSGSTHGTSLEKHKILHRLLODSSPYDLAKLTAEATG 712
      11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
      5 esssvtspgsvsstsgvstsmhbgslqekhrilhkgnspaeavakiteatg 64
      .

```

```

Qy      713 KDLSSSSSTAPSGEVTIKQEPVSPKKKE-NALLRYLLDKDDTKDIGLPETPKLERLDS 771
      11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
      65 kd--tsstlscgdgnvqvqeqispkpkennallryllldrdpsdalskelqpqvadvn 121
Db
Qy      772 KTDPAANTKLIAMKTEKE 789
      11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
      122 kmsqctsstipsssgqkd 139
Db
RESULT  33
ABB21964
ID      ABB21964 standard; Protein; 149 AA.
XX
AC      ABB21964;
XX
DT      23-JAN-2002 (first entry)
XX
DE      Protein #3963 encoded by probe for measuring heart cell gene expression.
XX
DE      Protein #3963 encoded by probe for measuring heart cell gene expression.
XX
KW      Human; gene expression; heart; microarray; vascular system;
XX
KW      cardiovascular disease; hypertension; cardiac arrhythmia;
XX
KW      congenital heart disease.
XX
OS      Homo sapiens.
XX
PN      WO200157274-A2.
XX
PD      09-AUG-2001.
XX
PF      30-JAN-2001; 2001WO-US00666.
XX
PR      04-FEB-2000; 2000US-0180312.
XX
PR      26-MAY-2000; 2000US-0207456.
XX
PR      30-JUN-2000; 2000US-0608408.
XX
PR      03-AUG-2000; 2000US-0632366.
XX
PR      21-SEP-2000; 2000US-0234687.
XX
PR      27-SEP-2000; 2000US-0236359.
XX
PR      04-OCT-2000; 2000GB-0024263.
XX
PA      (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI      Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR      WPI; 2001-488899/53.
XX
PT      Single exon nucleic acid probes for analyzing gene expression in human
XX
PT      hearts -
XX
PS      Claim 15; SEQ ID NO 23734; 530pp; English.
XX
CC      The present invention relates to single exon nucleic acid probes for
XX
CC      measuring human gene expression in a sample derived from human heart (see
XX
CC      ABA21535-ABA41305). The present sequence is a protein encoded by one such
XX
CC      probe. The probes may be used for predicting, measuring and displaying
XX
CC      gene expression in samples derived from the human heart via microarrays.
XX
CC      By measuring gene expression, the probes are useful for predicting,
XX
CC      diagnosis, grading, staging, monitoring and prognosing diseases of the
XX
CC      human heart and vascular system e.g. cardiovascular disease.
XX
CC      hypertension, cardiac arrhythmias and congenital heart disease.
XX
CC      Note: The sequence data for this patent did not form part of the printed
XX
CC      specification, but was obtained in electronic format directly from WIPO
XX
CC      at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ      Sequence 149 AA:

```

```

Query Match      3.8%; Score 287; DB 22; Length 149;
Best Local Similarity 44.2%; Pred. No. 4.4e-11;
Matches 61; Conservative 31; Mismatches 42; Indels 4; Gaps 2;

```

```

Qy      653 EPSPLASSLSDTNKDGSLPGSGSTHGTSLEKHKILHRLLODSSPYDLAKLTAEATG 712
      11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11

```

```

Db      5  essvsvtspgsvsstsgvstsmhgsllqekhrllhkl1lqngnspeavakltaeatg 64
Qy      713 KDLQSSSTFADPGSEVTKQEPVSPKKE-NALLRYLLDKDDTKDIGLPEITPKLERLDS 771
        || || | : | : ||| : ||||| ||||| ||||| ||| | : | : | : | : | :
Db      65 kd---tsitscggdgvnvkqglspkkennallryllldrdpsdalskelqpvegydn 121
Qy      772 KTDPAASNTKLIAMKTEKE 789
        | : : : | : | :
Db      122 kmsqctstlpsseqkd 139

RESULT 34
AAM57393
ID      AAM57393 standard; Protein: 149 AA.
XX
XX
AC      AAM57393;
XX
DT      05-NOV-2001 (first entry)
XX
DE      Human brain expressed single exon probe encoded protein SEQ ID NO: 29498.
XX
KW      Human: brain expressed exon; gene expression analysis; probe;
KW      microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW      epilepsy; cancer.
XX
OS      Homo sapiens.
XX
PN      WO200157275-A2.
XX
PD      09-AUG-2001.
XX
PE      30-JAN-2001; 2001WO-US00667.
XX
PR      04-FEB-2000; 2000US-0180312.
PR      26-MAY-2000; 2000US-0207456.
PR      30-JUN-2000; 2000US-0608408.
PR      03-AUG-2000; 2000US-0632366.
PR      21-SEP-2000; 2000US-0234687.
PR      27-SEP-2000; 2000US-0236359.
PR      04-OCT-2000; 2000GB-0024263.
XX
PA      (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI      Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR      WPI: 2001-483446/52.
XX
PT      Single exon nucleic acid probes for analyzing gene expression in human
PT      brains -
XX
PS      Example 4; SEQ ID NO: 29498; 650pp + Sequence Listing; English.
XX
CC      The present invention provides a number of single exon nucleic acid
CC      probes which are derived from genomic sequences expressed in the human
CC      brain. They can be used to measure gene expression in brain cell samples,
CC      CC which may enable the diagnosis and improved treatment of nervous system
CC      diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC      epilepsy and cancers. The present sequence is a protein encoded by one of
CC      the probes of the invention.
XX
SQ      Sequence 149 AA;

Query Match 3.88; Score 287; DB 22; Length 149;
Best Local Similarity 44.28; Pred. No. 4.4e-11;
Matches 61; Conservative 31; Mismatches 42; Indels 4; Gaps 2;

Qy      653 EPSPLASSLSDTNKDSGTSLPGSGSTHGTSLKEKHILHRLLODSSPVDLAKLTAAATG 712
        | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db      5  essvsvtspgsvsstsgvstsmhgsllqekhrllhkl1lqngnspeavakltaeatg 64
Qy      713 KDLQSSSTFADPGSEVTKQEPVSPKKE-NALLRYLLDKDDTKDIGLPEITPKLERLDS 771
        || || | : | : ||| : ||||| ||||| ||||| ||| | : | : | : | : | :
Db      122 kmsqctstlpsseqkd 139

```

```

Db      65 kd---tsitscggdgvnvkqglspkkennallryllldrdpsdalskelqpvegydn 121
Qy      772 KTDPAASNTKLIAMKTEKE 789
        | : : : | : | :
Db      122 kmsqctstlpsseqkd 139

RESULT 35
AAM69785
ID      AAM69785 standard; Protein: 149 AA.
XX
XX
AC      AAM69785;
XX
DT      06-NOV-2001 (first entry)
XX
DE      Human bone marrow expressed probe encoded protein SEQ ID NO: 30091.
XX
KW      Human; bone marrow expressed exon; gene expression analysis; probe;
KW      microarray; cancer; leukaemia; lymphoma; myeloma.
XX
OS      Homo sapiens.
XX
PN      WO200157276-A2.
XX
PD      09-AUG-2001.
XX
PE      30-JAN-2001; 2001WO-US00668.
XX
PR      04-FEB-2000; 2000US-0180312.
PR      26-MAY-2000; 2000US-0207456.
PR      30-JUN-2000; 2000US-0608408.
PR      03-AUG-2000; 2000US-0632366.
PR      21-SEP-2000; 2000US-0234687.
PR      27-SEP-2000; 2000US-0236359.
PR      04-OCT-2000; 2000GB-0024263.
XX
PA      (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI      Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR      WPI: 2001-488900/53.
XX
PT      Human genome-derived single exon nucleic acid probes useful for
PT      analyzing gene expression in human bone marrow -
XX
PS      Example 4; SEQ ID NO: 30091; 658bp + Sequence Listing; English.
XX
CC      The present invention provides a number of single exon nucleic acid
CC      probes which are derived from genomic sequences expressed in the human
CC      bone marrow. They can be used to measure gene expression in bone marrow
CC      samples, which may enable the improved diagnosis and treatment of cancers
CC      such as lymphoma, leukemia and myeloma. The present sequence is a
CC      protein encoded by one of the probes of the invention.
XX
SQ      Sequence 149 AA;

Query Match 3.88; Score 287; DB 22; Length 149;
Best Local Similarity 44.28; Pred. No. 4.4e-11;
Matches 61; Conservative 31; Mismatches 42; Indels 4; Gaps 2;

Qy      653 EPSPLASSLSDTNKDSGTSLPGSGSTHGTSLKEKHILHRLLODSSPVDLAKLTAAATG 712
        | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db      5  essvsvtspgsvsstsgvstsmhgsllqekhrllhkl1lqngnspeavakltaeatg 64
Qy      713 KDLQSSSTFADPGSEVTKQEPVSPKKE-NALLRYLLDKDDTKDIGLPEITPKLERLDS 771
        || || | : | : ||| : ||||| ||||| ||||| ||| | : | : | : | : | :
Db      65 kd---tsitscggdgvnvkqglspkkennallryllldrdpsdalskelqpvegydn 121
Qy      772 KTDPAASNTKLIAMKTEKE 789
        | : : : | : | :
Db      122 kmsqctstlpsseqkd 139

```

## RESULT 36

AAM17609

ID AAM17609 standard; Protein; 149 AA.

AC AAM17609;

DT 12-OCT-2001 (first entry)

DE Peptide #4043 encoded by probe for measuring cervical gene expression.

KW Probe: human; microarray; gene expression; cervical epithelial cell;

KM cervical cancer.

XX Homo sapiens.

XX MO200157278-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00670.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488901/53.

XX Human genome-derived single exon nucleic acid probes useful for

XX analyzing gene expression in human cervical epithelial cells -

XX Claim 27; SEQ ID No 22435; 487pp; English.

XX The present invention relates to human single exon nucleic acid probes

XX (SENP: see AAI10068-AI128459). The present sequence is a peptide encoded

XX by one such probe. The SENPs are derived from human HeLa cells. The SENPs

XX can be used to produce a single exon microarray, which can be used for

XX measuring human gene expression in a sample derived from human cervical

XX epithelial cells. By measuring gene expression, the probes are therefore

XX useful in grading and/or staging of diseases of the cervix, notably

XX cervical cancer.

XX Note: The sequence data for this patent did not form part of the printed

XX specification, but was obtained in electronic format directly from WIPO

XX at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 149 AA:

SQ

Query Match 3.8%; Score 287; DB 22; Length 149;  
Best Local Similarity 44.2%; Pred. No. 4.4e-11;  
Matches 61; Conservative 31; Mismatches 42; Indels 4; Gaps 2;

QY 653 EPSPPLASSLSDTNKDSGSLPGSGSTHTGSLKSKKHTLHLDDSSSPVDLAKTLTAATG 712

DB 5 ESSSVYTSIPSGVSSSTGVSSTSMHGSILQEKHRIHLILQNGSPAEVAKITAEATG 64

QY 713 KDLSOESSSTAPGSEVTVIKOEPVSPKKE-NALLRVLDDKDPDKDIGLPETTPKLERLDS 771

DB 65 Kd---tsstscgdgnvvkqegispkkennallrylltdcpdsalskelqpvegvdn 121

QY 772 KTDPAANTKLIAKTEKE 789

DB 122 kmsqctsstipsssgkd 139

## RESULT 37

AAM30127

ID AAM30127 standard; Protein; 149 AA.

AC AAM30127;

DT 17-OCT-2001 (first entry)

DE Peptide #4164 encoded by probe for measuring placental gene expression.

KW Probe: microarray; human; placenta; antenatal diagnosis;

KM genetic disorder.

XX Homo sapiens.

XX MO200157272-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00663.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488897/53.

XX Human genome-derived single exon nucleic acid probes useful for

XX analyzing gene expression in human placenta -

XX Claim 27; SEQ ID No 30396; 654pp; English.

XX The present invention relates to single exon nucleic acid probes (SENP:

XX see AAI13315-AI157546). The present sequence is a peptide encoded by one

XX such probe. The probes are useful for producing a microarray for

XX predicting, measuring and displaying gene expression in samples derived

XX from human placenta. The probes are useful for antenatal diagnosis of

XX human genetic disorders.

XX Sequence 149 AA:

SQ

Query Match 3.8%; Score 287; DB 22; Length 149;  
Best Local Similarity 44.2%; Pred. No. 4.4e-11;  
Matches 61; Conservative 31; Mismatches 42; Indels 4; Gaps 2;

QY 653 EPSPPLASSLSDTNKDSGSLPGSGSTHTGSLKSKKHTLHLDDSSSPVDLAKTLTAATG 712

DB 5 ESSSVYTSIPSGVSSSTGVSSTSMHGSILQEKHRIHLILQNGSPAEVAKITAEATG 64

QY 713 KDLSOESSSTAPGSEVTVIKOEPVSPKKE-NALLRVLDDKDPDKDIGLPETTPKLERLDS 771

DB 65 Kd---tsstscgdgnvvkqegispkkennallrylltdcpdsalskelqpvegvdn 121

QY 772 KTDPAANTKLIAKTEKE 789

DB 122 kmsqctsstipsssgkd 139

## RESULT 38

AAM05268

ID AAM05268 standard; Protein; 149 AA.

AC AAM05268;



DT 09-OCT-2001 (first entry)  
XX Peptide #3950 encoded by probe for measuring breast gene expression.  
DE  
XX Probe: human; breast disease; breast cancer; development disorder;  
KW Inflammatory disease; proliferative breast disease; non-carcinoma tumour.  
XX Homo sapiens.  
OS  
XX WO200157270-A2.  
PN  
XX 09-AUG-2001.  
PD  
XX 29-JAN-2001; 2001WO-US00661.  
PF  
XX 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
PR  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
PA Penn SG, Hanzel DK, Chen W, Rank DR;  
PI WPI; 2001-476286/51.  
XX  
DR Novel single exon nucleic acid probe used to measuring gene expression  
XX in a human breast -  
XX  
XX Claim 27; SEQ ID NO 14008; 322pp; English.  
XX  
CC The present invention relates to novel single exon nucleic acid probes  
CC (see AAI00010-AA110067). The present sequence is a peptide encoded by one  
CC such probe. The probes are useful for measuring human gene expression in  
CC a human breast sample, where the probe hybridises at high stringency to a  
CC nucleic acid expressed in the human breast. The probes are useful for  
CC predicting, diagnosing, grading, staging, monitoring and prognosing  
CC diseases of the human breast, particularly those diseases with polygenic  
CC aetiology. The diseases include: breast cancer, disorders of development,  
CC inflammatory diseases of the breast, fibrocystic changes, proliferative  
CC breast disease and non-carcinoma tumours.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 149 AA;  
SQ  
Query Match 3.8%; Score 287; DB 22; Length 149;  
Best Local Similarity 44.2%; Pred. No. 4.4e-11;  
Matches 61; Conservative 31; Mismatches 42; Indels 4; Gaps 2;  
QY 653 EPSPLASSISDYNKSTGSLPGSGSTHCTSLKEKHILRILODSSPVDIAKLTAATG 712  
Db 5 ESSVSVTSPEVSSTSGVSTSTNMHSLIGKHRIHLKILQNGSPAFAAKIAAATG 64  
QY 713 KILSQESSSTAPGSEVTIKQEPVSPKKE-NALLRYLLDKDTRDIGLPETPKLERADS 771  
Db 65 kd--tsstiscgdnvkvqglapkkkenallrlyllrdpdsalskeiqvgydn 121  
QY 772 KTDPASTKILAMTEKE 789  
Db 122 kmsqctslipssqekd 139

RESULT 39  
AAU16186  
ID AAU16186 standard; Protein; 591 AA.  
XX  
AC AAU16186;

XX 07-NOV-2001 (first entry)  
DT Human novel secreted protein, Seq ID 1139.  
XX  
DE  
XX Human; immunosuppressive; antiarthritic; antirheumatic;  
KW cytosolatic; cardiant; vasotropic; cerebroprotective; nootropic;  
KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;  
KW vulnerrary; secreted protein; rheumatoid arthritis;  
KW hyperproliferative disorder; cardiovascular disorder; cardiac arrest;  
KW cerebrovascular disorder; cerebral ischaemia; angiogenesis;  
KW nervous system disorder; Alzheimer's disease; infection; ocular disorder;  
KW corneal infection; wound healing; epithelial cell proliferation;  
KW skin ageing; food additive; preservative; antiproliferative.  
XX  
XX Homo sapiens.  
OS  
XX WO200155322-A2.  
PN  
XX 02-AUG-2001.  
PD  
XX 17-JAN-2001; 2001WO-US01341.  
PF  
XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.

PR	18-SEP-2000	2000US-0233068
PR	12-SEP-2000	2000US-0233061
PR	14-SEP-2000	2000US-0233397
PR	14-SEP-2000	2000US-0233398
PR	14-SEP-2000	2000US-0233399
PR	14-SEP-2000	2000US-0233400
PR	14-SEP-2000	2000US-0233401
PR	14-SEP-2000	2000US-0233063
PR	14-SEP-2000	2000US-0233064
PR	14-SEP-2000	2000US-0233065
PR	21-SEP-2000	2000US-0234122
PR	21-SEP-2000	2000US-0234123
PR	25-SEP-2000	2000US-0234997
PR	25-SEP-2000	2000US-0234998
PR	25-SEP-2000	2000US-0234999
PR	29-SEP-2000	2000US-0236366
PR	29-SEP-2000	2000US-0236367
PR	29-SEP-2000	2000US-0236368
PR	29-SEP-2000	2000US-0236370
PR	02-OCT-2000	2000US-0236802
PR	02-OCT-2000	2000US-0237037
PR	02-OCT-2000	2000US-0237038
PR	02-OCT-2000	2000US-0237039
PR	13-OCT-2000	2000US-0237040
PR	13-OCT-2000	2000US-0239355
PR	13-OCT-2000	2000US-0239357
PR	20-OCT-2000	2000US-0240960
PR	20-OCT-2000	2000US-0241121
PR	20-OCT-2000	2000US-0241785
PR	20-OCT-2000	2000US-0241786
PR	20-OCT-2000	2000US-0241787
PR	20-OCT-2000	2000US-0241808
PR	20-OCT-2000	2000US-0241809
PR	01-NOV-2000	2000US-0241826
PR	01-NOV-2000	2000US-0246117
PR	08-NOV-2000	2000US-0246474
PR	08-NOV-2000	2000US-0246475
PR	08-NOV-2000	2000US-0246476
PR	08-NOV-2000	2000US-0246477
PR	08-NOV-2000	2000US-0246478
PR	08-NOV-2000	2000US-0246523
PR	08-NOV-2000	2000US-0246524
PR	08-NOV-2000	2000US-0246525
PR	08-NOV-2000	2000US-0246526
PR	08-NOV-2000	2000US-0246527
PR	08-NOV-2000	2000US-0246528
PR	08-NOV-2000	2000US-0246532
PR	08-NOV-2000	2000US-0246609
PR	08-NOV-2000	2000US-0246610
PR	08-NOV-2000	2000US-0246611
PR	08-NOV-2000	2000US-0246613
PR	17-NOV-2000	2000US-0249207
PR	17-NOV-2000	2000US-0249208
PR	17-NOV-2000	2000US-0249209
PR	17-NOV-2000	2000US-0249210
PR	17-NOV-2000	2000US-0249211
PR	17-NOV-2000	2000US-0249212
PR	17-NOV-2000	2000US-0249213
PR	17-NOV-2000	2000US-0249214
PR	17-NOV-2000	2000US-0249245
PR	17-NOV-2000	2000US-0249264
PR	17-NOV-2000	2000US-0249265
PR	17-NOV-2000	2000US-0249297
PR	17-NOV-2000	2000US-0249299
PR	17-NOV-2000	2000US-0249300

PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250392.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251989.  
PR 11-DEC-2000; 2000US-0251990.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
PI  
PI Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI; 2001-488783/53.  
DR N-PSDB; AAS26173.  
XX  
XX  
PT New nucleic acid molecules encoding 461 human secreted proteins for  
PT diagnosing, preventing, treating or ameliorating medical conditions and  
PT used as food additives or preservatives -  
XX  
XX  
PS Claim 11; SEQ ID No 1139; 980BP; English.

The invention relates to isolated nucleic acid molecules and their encoded secreted proteins. The nucleic acids and proteins are used to prevent, treat or ameliorate a medical condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used in diagnosing a pathological condition or susceptibility to a pathological condition. Antibodies to the proteins can also be used in alleviating symptoms associated with the disorders and in diagnostic immunoassays e.g. radioimmunoassays or enzyme linked immunosorbent assays (ELISA). Disorders which are diagnosed or treated include autoimmune diseases e.g. Rheumatoid arthritis, hyperproliferative disorders e.g. neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g. Alzheimer's disease, infections caused by bacteria, viruses and fungi and ocular disorders e.g. corneal infection, and many other disorders listed in the specification. The polypeptides can also be used to aid wound healing and epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain organs before transplantation, for supporting cell culture of primary tissues, to regenerate tissues and in chemotaxis. The polypeptides can also be used as a food additive or preservative to increase or decrease storage capabilities, fat content, lipid, protein, carbohydrate, vitamins, minerals, cofactors and other nutritional components. The present sequence represents a novel secreted protein of the invention.

Query	Match	3.7%	Score	285	DB	22	Length	591
Best Local	Similarity	22.0%	Pred	No. 5.1e-10				
Matches	130	Conservative	94	Mismatches	250	Indels	118	Gaps
								22.
QY	17	RRKKECPDGLGSPKRNTEKRNREQENKYYIEELAEILFANFNDIDNFNPKCALKET	76					
Db	53	rkrqdsdsgsqeansqtekrtrrcknmnlleelsam-----pqcpmarklklvllma	108					
QY	77	VOQIQIKOEKAAAAANIDEVOKSDVSSRGQGVIDKDALGPMMLLELDQFFRVNLE--CN	135					
Db	109	vghltslkgltusvyg-----snypsfldqnelhllkvaegflitvycerjk	158					
QY	136	VEFVSENVQYRYNOELMNKSVYSILHVGHTEFEVKMLLPKSLVNG-----GSMSCP	190					
Db	159	llfveksvskllyngdastlqsglfdflhpkdvak-vkeqslsfdslspeklldaktgyl	217					
QY	191	PKRNSH-----TFNCRMALVPLPDSDEES--HDNOEAKQKETMOC--FANS	233					
Db	218	vshnlagrltyvysrrtsfcrllkscklsvkeegncjpsnkckekrrkylthlctgyls	277					
QY	234	QPKSIKEBEDQS-----CLICVARR--VPMKERPALPSESFTTRQDLOGKIT	281					





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 7, 2002, 10:35:38 ; Search time 31.5 Seconds  
(without alignments)  
4465.866 Million cell updates/sec

Title: US-09-842-256-2

Perfect score: 7631

Sequence: 1 MSGMGENTSDPSRAETRRKRK.....NQLPGMDMKQEGDTTRKYC 1464

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR\_71:\*  
2: pir1:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7216	94.6	1462	2 T42639	glucocorticoid rec
2	6659.5	87.3	1463	2 T30193	nuclear receptor c
3	2729	35.8	1424	2 T03851	thyroid hormone re
4	1033.5	13.5	1061	2 A57620	steroid receptor c
5	340	4.5	2232	2 T34434	hypothetical prote
6	319	4.2	1505	2 JC4851	hypoxia-inducible
7	313.5	4.1	4957	2 T03455	AR protein - huma
8	313.5	4.1	5262	2 T03454	AR protein - huma
9	309	4.0	2715	2 T13049	eyelid - fruit fly
10	297.5	3.9	1088	2 H96747	unknown protein T1
11	292	3.8	2649	2 T51023	hypothetical prote
12	287	3.8	1234	2 T30160	hypothetical prote
13	282	3.7	3498	2 T22330	hypothetical prote
14	280	3.7	925	2 T19361	hypothetical prote
15	277	3.6	2414	2 A54277	transcription adap
16	275	3.6	2722	2 T20532	hypothetical prote
17	273.5	3.6	1366	2 B86292	FTH2.12 protein -
18	272	3.6	626	2 JE0270	Arnt-like PAS prot
19	270.5	3.5	2738	2 E88320	protein F07A11.6 l
20	270	3.5	667	2 JC7771	hypoxia inducible
21	270	3.5	1589	2 T13606	hypothetical prote
22	269.5	3.5	826	2 T138972	hypoxia-inducible
23	269	3.5	626	2 JC5405	brain and muscle A
24	269	3.5	1212	2 T13804	she protein - frui
25	269	3.5	1952	2 T48814	hypothetical prote
26	268.5	3.5	1058	2 T30556	aryl hydrocarbon r
27	268	3.5	1142	2 T00022	B120 protein - hum
28	266.5	3.5	2090	2 T52058	probable transform
29	264.5	3.5	1307	2 T25563	hypothetical prote

30	263.5	3.5	1059	2 T30557	aryl hydrocarbon r
31	261.5	3.4	2282	2 T42717	DNA-binding protei
32	258.5	3.4	3507	2 T34513	hypothetical prote
33	258	3.4	2215	2 T16871	hypothetical prote
34	257.5	3.4	1655	2 T13998	hypothetical prote
35	255.5	3.3	813	2 JC5809	gene mastermind pr
36	254.5	3.3	2606	2 T24157	hypoxia-inducible
37	253.5	3.3	2584	2 T24158	hypothetical prote
38	252	3.3	1819	2 T32008	hypothetical prote
39	252	3.3	2441	2 S39161	hypothetical prote
40	252	3.3	3190	2 T13828	CREB-binding prote
41	251	3.3	1794	2 T38459	CREB-binding prote
42	250	3.3	1023	2 T13068	hypothetical diver
43	250	3.3	1596	2 A33106	CLOCK protein - fir
44	248.5	3.3	3968	2 A44265	neurogenic locus m
45	248	3.2	2175	1 S03170	trithorax homolog homeotic protein c

#### ALIGNMENTS

RESULT	1
T42639	
glucocorticoid receptor interacting protein GRP1 - mouse	
C:Species: Mus musculus (house mouse)	
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 11-Jan-2000	
C:Accession: T42639	
R:Hong, H.; Kohli, K.; Garabedian, M.J.; Stallcup, M.R.	
Mol. Cell. Biol. 17, 2735-2744, 1997	
A>Title: GRP1, a transcriptional coactivator for the AF-2 transactivation domain of	
A:Reference number: Z22229; MUID:97265407	
A:Accession: T42639	
A>Status: preliminary; translated from GB/EMBL/DBJ	
A:Molecule type: mRNA	
A:Residues: 1-1462 <HON>	
A:Cross-references: EMBL:U39060; NID:G1853979; PID:G1853980; PIDN:AAC53151.1	
C:Genetics:	
A:Gene: GRP1	
C:Function:	
A:Description: transcriptional co-activator for steroid receptors and nuclear recepto	
A>Note: AF-2 requires GRP1 as a coactivator, but AF-1 does not	

  

Query Match	94.6%	Score 7216;	DB 2;	Length 1462;
Best Local Similarity	94.1%	Pred. No. 0;		
Matches 1377;	Conservative 43;	Mismatches 42;	Indels 2;	Gaps 1;

  

QY	1	MSGMGENTSDPSRAETRRKKECPDOLGSPKRNTERKRNREQENKYIEELAEILIFANFNDI	60
DB	1	MSGMGENTSDPSRAETRRKKECPDOLGSPKRNTERKRNREQENKYIEELADLIFANFNDI	60
QY	61	DNFNRPKDKALIKETVQKQIRKQEKRAAANIDEVOKSDVSSFGGCVIDKDALGPMML	120
DB	61	DNFNKPKDKALIKETVQKQIRKQEKRAAANIDEVOKSDVSSFGGCVIDKDALGPMML	120
QY	121	EALDGFPPVNLKGNVVPSEENVTOYLRYNOBELNKKSVSYLHGHDTFEPKNNLPKSI	180
DB	121	EALDGFPPVNLKGNVVPSEENVTOYLRYNOBELNKKSVSYLHGHDTFEPKNNLPKSM	180
QY	181	VNGSGMSGEPFRNRSHFTNCRLVPLPDSEEGHDNOEAHQKYETMOCFAVSOQSIKE	240
DB	181	VNGSGMSGEPFRNRSHFTNCRLVPLPDSEEGHDNOEAHQKYETMOCFAVSOQSIKE	240
QY	241	EGEDLOSLICVARRVPKKEKRPVLPSSSFTTRDLOGKITSLDSTYRAAMKPGMEDLV	300
DB	241	EGEDLOSLICVARRVPKKEKRPVLPSSSFTTRDLOGKITSLDSTYRAAMKPGMEDLV	300
QY	301	RRCIOKPAHNGHGSVSAKRRHNEVLROGLAFSQTYPFSSDGLVAQAORSKYIRISOT	360
DB	301	RRCIOKPAHNGHGSVSAKRRHNEVLROGLAFSQTYPFSSDGLVAQAORSKYIRISOT	360
QY	361	TNEPOLVLSLHMLHREQVNCVNPDLTGOTMGKPLNPITSSNSPAHQALCSGPNQGDMLTS	420
DB	361	TNEPOLVLSLHMLHREQVNCVNPDLTGOTMGKPLNPITSSNSPAHQALCSGPNQGDMLTS	420

```

Db 361 TNEPQVYISLHMLHREQNVCMNPDLTGQAMGKPLNPISSSPAHQALCSGQNPQDWTLG 420
QY 421 SNINFPINGKEQMGMPGREGSGGMNHYSGMOATTPOGSNNVALKMNNSPQSSPGMNP 480
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 421 SNINFPINGKEQMGMPGREGSGGMNHYSGMOATTPOGSNNVALKMNNSPQSSPGMNP 480
QY 481 QPMSLSPRHRMSGVAGSPRIPPSQSPAGSLSPVGVCSSTGNSHTNSSTNLALQAL 540
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 481 QASVYLSPRHRMSGVAGSPRIPPSQSPAGSLSPVGVCSSTGNSHTNSSTNLALQAL 540
QY 541 SEGHGVSLSGLSLASPDLMKGNLONSPVNMNPPPLSKMGLSDSKDCFGLYGPSPGTTGOA 600
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 541 SEGHGVSLSGLSLASPDLMKGNLONSPVNMNPPPLSKMGLSDSKDCFGLYGPSPGTTGOA 600
QY 601 ESSCHPEQKEETNDPNLPVAVSERADGSRSLHDSKQOTKLLQLLTTRKSDQMEPSPLASS 660
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 601 EASCHPEQKPNDSMPQAASGDRAGHSRLHDSKQOTKLLQLLTTRKSDQMEPSPLASS 660
QY 661 LSDTNKDSSTGSLPGSGSTHGTSLKEKHKILHRLLDSSSPVDLAKLTAEATGKDLQESS 720
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 661 LSDTNKDSSTGSLPGSGSTHGTSLKEKHKILHRLLDSSSPVDLAKLTAEATGKDLQESS 720
QY 721 STAPGSEVTIKQEPVSKKKNALLRYLLDKDPTKDIGLPETPKLERLDSKTDPAANTK 780
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 721 STAPGSEVTIKQEPVSKKKNALLRYLLDKDPTKDIGLPETPKLERLDSKTDPAANTK 780
QY 781 LIAMKTEKEEMSEEPDQPGSELNDEIILLDLQNSQLPOLFPDTRPAGABAVDQKAI 840
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 781 LIAMKTEKEEVSFEPSDQPGSELNDEIILLDLQNSQLPOLFPDTRPAGABAVDQKAI 840
QY 841 NDLMQTLAENSPTVYVCAQKTALRISQSTFNPNRPGQLGRLPNQNLPLDITLQSPGAG 900
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 841 NDLMQTLADSSVPVPAQAQALMSQSTFNPNRPGQLGRLPNQNLPLDITLQSPGAG 900
QY 901 PEPPIRNSSTSYVLPQPGMNGNOGMINQCNLGSSTGMIGNSARPTMPSGEMAPQSSA 960
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 901 PEPPIRNSSTSYVLPQPGMNGNOGMINQCNLGSSTGMIGNSARPTMPSGEMAPQSSA 960
QY 961 VRVTCATTSAMNRPVQGMIRNPAASIPMRPSSQPGQROTLOQVYNNIGPSELEMMNG 1020
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 961 VRVTCATTSAMNRPVQGMIRNPAASIPMRPSSQPGQROTLOQVYNNIGPSELEMMNG 1020
QY 1021 PVSQQAAPRNOTAPWPESTLPTDQASFASQONQPPSSSDDLCPHRAEESPDECALL 1080
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1021 PVSQQAAPRNOTAPWPESTLPTDQASFASQONQPPSSSDDLCPHRAEESPDECALL 1080
QY 1081 DOYLALRNFDGLLEIDRALGIPELVYSQAVDAEOPSSQESSIMLEQKPPVFPQYASQ 1140
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1081 DOYLALRNFDGLLEIDRALGIPELVYSQAVDAEOPSSQESSIMLEQKPPVFPQYASQ 1140
QY 1141 AQMAQGSYSFMDQPNFHTMGQRPSTATLRMQPRPGLRPTGLVQNPQRLQLOHRLQAO 1200
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1141 AQMAQGSYSFMDQPNFHTMGQRPSTATLRMQPRPGLRPTGLVQNPQRLQLOHRLQAO 1200
QY 1201 QNTQPLMNOISNVSNNVLTLRGVPYTPQAPINAQMLAORQREILLNOHRLQOM 1260
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1201 QNTQPLMNOISNVSNNVLTLRGVPYTPQAPINAQMLAORQREILLNOHRLQOM 1260
QY 1261 RTLLMNRQGLNMTPSVYAPSPATMSNPRIPQANAOQFPFPYNYGISOQPDGFTGATT 1320
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1261 RTLLMNRQGLNMTPSVYAPSPATMSNPRIPQANAOQFPFPYNYGISOQPDGFTGATT 1320
QY 1321 POSPLMSPRAHQPSPMAQSOANPAVQAPSDINGMAGNMGNMSFQSQSPHFGQAN 1380
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1321 POSPLMSPRAHQPSPMAQSOANPAVQAPSDINGMAGNMGNMSFQSQSPHFGQAN 1380
QY 1381 TSMTSNMNTINVSMAITNGMSSNMONTGOISMTSVTSVSTGLSSKGPQVNDPALRG 1440
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1381 TSMTSNMNTINVSMAITNGMSSNMONTGOISMTSVTSVSTGLSSKGPQVNDPALRG 1440
QY 1441 NLFPNOLPGMDMTKQGBDPTTRKYC 1464
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1441 NLFPNOLPGMDMTKQGBDPTTRKYC 1464
QY 1439 NLFPNOLPGMDMTKQGBDASRRKYC 1462
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

RESULT 2
T30193
nuclear receptor coactivator protein 2 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 22-Oct-1999 #sequence_rev15ion 22-Oct-1999 #extL_change 22-Oct-1999
R:Torchia, J.; Rose, D.W.; Inostroza, J.; Kamel, Y.; Westlin, S.; Glass, C.K.; Rosenfe
Nature 387, 677-684, 1997
A:Title: The transcriptional co-activator p/CIP binds CBP and mediates nuclear-recept
A:Reference number: Z20768; M0ID:97336097
A:Accession: T30193
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1463 <TOR>
A:Cross-references: EMBL:AF000582; NID:92213816; PTD:92213817; PIDN:AA61575.1

Query Match 87.3%; Score 6659.5; DB 2; Length 1463;
Best Local Similarity 87.6%; Pred. No. 0;
Matches 1298; Conservative 60; Mismatches 87; Indels 37; Gaps 10;

QY 1 MSGGENTSPPSAETRRKRECPDQLGSPKRNTERKRNQENKYIELELIFANFDI 60
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MSGGENTSPPSAETRRKRECPDQLGSPKRNTERKRNQENKYIELELIFANFDI 60
QY 61 DNFNFKPKCAILKETVQJROIKEQEKAAANIDEVQKSDVSTGGVIDKALGPMML 120
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 DNFNFKPKCAILKETVQJROIKEQEKAAANIDEVQKSDVSTGGVIDKALGPMML 120
QY 121 EALDGEFFVNLGQNVVSENVTOYLRYNOBELMKNKSVYSLHVGDTHEFVKNLLPKST 180
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 EALDGEFFVNLGQNVVSENVTOYLRYNOBELMKNKSVYSLHVGDTHEFVKNLLPKSM 180
QY 181 VNGGSMGGEPRRNSHTFNRMLVYKLPDSEBEGHNOEHQKTYETMQCAVQPKSIKE 240
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 181 VNGGSMGGEPRRNSHTFNRMLVYKLPDSEBEGHNSQEHQKTYEAMQCAVQPKSIKE 240
QY 241 EGEDLOGLCLVARYVYMKR-----PVLSSSEFTTRDOLQKITSLDPTSM 288
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 241 EGEDLOGLCLVARYVYMKR-----PVLSSSEFTTRDOLQKITSLDPTSM 288
QY 289 RAAM-KPWEDLVKRCIQKFHQAQHEGESYAKRHHHEVLROGLAFSQIYRFLSDGTV 347
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 289 RAAM-KPWEDLVKRCIQKFHQAQHEGESYAKRHHHEVLROGLAFSQIYRFLSDGTV 347
QY 348 AAQTKSLIRSQTTNEPQVYISLHMLHREQNVCMNPDLTGQMGKPLNPISSSPAHQA 407
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 348 AAQTKSLIRSQTTNEPQVYISLHMLHREQNVCMNPDLTGQMGKPLNPISSSPAHQA 407
QY 408 LCGSNPGQDMTTLSSNINFPINGKEQMGMPGREGSGGMNHYSGMOATTPOGSNNALKA 467
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 408 LCGSNPGQDMTTLSSNINFPINGKEQMGMPGREGSGGMNHYSGMOATTPOGSNNALKA 467
QY 468 NSPQSSPGMNPQPTSMLSPRHRMSPVAGSPRIPPSQSPAGSLSPVGVCSSTGNSH 527
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 468 NSPQSSPGMNPQPTSMLSPRHRMSPVAGSPRIPPSQSPAGSLSPVGVCSSTGNSH 527
QY 528 STYNSSTNLALQALSEGHSVLSGLSLASPDLMKGNLONSPVNMNPPPLSKMGLSDSK 587
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 528 STYNSSTNLALQALSEGHSVLSGLSLASPDLMKGNLONSPVNMNPPPLSKMGLSDSK 587
QY 588 LYGEPSGCTGQAESSCHPEQKETNDPNLPVAVSERADGSRSLHDSKQOTKLLQLLTT 647
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 588 LYGEPSGCTGQAESSCHPEQKETNDPNLPVAVSERADGSRSLHDSKQOTKLLQLLTT 647
QY 648 KSDQMEPSPLASSLSDTNKDSSTGSLPGSGSTHGTSLKEKHKILHRLLDSSSPVDLAKLT 707
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 648 KSDQMEPSPLASSLSDTNKDSSTGSLPGSGSTHGTSLKEKHKILHRLLDSSSPVDLAKLT 707
QY 708 AEATGKDLQESSSTAPGSEVTIKQEPVSKKKNALLRYLLDKDPTKDIGLPETPKLE 767
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

Db 708 AEATGKELSOESSSTAPEGSEVTVKOEAPSPKKENALLRYLLDKDXTKIDGILEPITPKLE 767  
 Qy 768 RLDSKTDPASTKLIAMTKEEMSEEPDQPGSELNDLEETLIDLOKSOI.POLEPDRP 827  
 Db 768 RLDSKTDPASTKLIAMTKEEVSEFSDPGSELNDLEETLIDLOKSOI.POLEPDRP 827  
 Qy 828 GAPAGSYDQKAIINDLMOLTAENSPVPEVGAOKTALRISOSTFNNPRGOLRLPNONTL 887  
 Db 828 GAPGYSVDKAIINDLMOLTAENSPVPEVGAOKTALRISOSTFNNPRGOLRLPNONTL 887  
 Qy 888 PLDITLOSPTAGPFPPIRNSSPVPIRQPMGNOGICNOGICNSTGMIGNSASRP 947  
 Db 888 PLDITLOSPTAGPFPPIRNSSPVPIRQPMGNOGICNOGICNSTGMIGNSASRP 947  
 Qy 948 TMRGEMAPSSAVRYTCAAT---TSAMNRPVGGMIR--NPASISMRPSSQPGQRTL 1002  
 Db 948 SMPGEMAPOST---SEESTLVLLPLVPRTDQSKEARFNPJTASIPGANSQGLQKML 1003  
 Qy 1003 QSOVMNIGPSELENNMGSPQYSOOOAPPNOTAPWPESTILPTDOASFASQNRQPGSSPD 1062  
 Db 1004 QSOVMNIGPSELENNMGSPQYNQOQAPPNOTAPWPESTILPTDOASFASQNRQPGSSPD 1063  
 Qy 1063 LLCFHPAAESPSDGCALIDLYLALRNFDELIEDRALGIDELVSQSOAVDEQFSODS 1122  
 Db 1064 LLCFHPAAESPSDGCALIDLYLALRNFDELIEDRALGIDELVSQSOAVDEQFSODS 1123  
 Qy 1123 NIMLEOKAPVPRQOYASQOAGOSYSPMDPNFHTMGORSPYATLRMOPRGLRPTGLV 1182  
 Db 1124 SIMLEOKAPVPRQOYASQOAGOSYSPMDPNFHTMGORSPYATLRMOPRGLRPTGLV 1183  
 Qy 1183 QNQPNOQLRLQHRILQAOQNROPPLMNOISVSNVNLTLRPGVPTQAPINAOMLAQOREI 1242  
 Db 1184 QNQPNOQLRLQHRILQAOQNROPPLMNOISVSNVNLTLRPGVPTQAPINAOMLAQOREI 1243  
 Qy 1243 LNOHLROROMHQOQOYQOQRTLMARGCGICNMTPSVAVSGMAYTNSNPRIPOANAOQFPFP 1302  
 Db 1244 LNOHLROROM---QOQOYQOQRTLMARGCGICNMTPSVAVSGMAYTNSNPRIPOANAOQFPFP 1301  
 Qy 1303 PNYGISOOPDGFPGATTPOSLMSPRAHTQSPMOOSQANPAYOAFSDINGNAQOMG 1362  
 Db 1302 PNYGISOOPDGFPGATTPOSLMSPRAHTQSPMOOSQANPAYOAFSDINGNAQOMG 1361  
 Qy 1363 GNSMFQOQSPHFGQOANTSMYSNNMNIIVSMATNTGSMSSNMOTGOISMTSVTSVS 1422  
 Db 1362 GNSMFQOQSPHFGQOANTSMYSNNMNIIVSMATNTGSMSSNMOTGOISMTSVTSVS 1421  
 Qy 1423 GLSSMGPEOVNDPALRGCNLFPPNOLPGMDATKOGDITTRKTC 1464  
 Db 1422 GLPSMGPEOVNDPALRGCNLFPPNOLPGMDATKOGDASRRKTC 1463

RESULT 3  
 T03851  
 thyroid hormone receptor activator molecule - human  
 C:Species: Homo sapiens (man)  
 C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 01-Dec-2000  
 C:Accession: T03851: T03749: T03443  
 R:Tekeshita, A.; Cardona, G.R.; Koibuchi, N.; Suen, C.S.; Chin, W.W.  
 J. Biol. Chem. 272, 27629-27634, 1997  
 A:Title: TRAM-1, a novel 160-kDa thyroid hormone receptor activator molecule, exhibits d  
 A:Reference number: 215120; M01D:98010595  
 A:Accession: T03851  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-1424 <TRAM>  
 A:Cross-references: EMBL:AF016031; NID:g2584879; PIDN:AA051849.1; PID:g2584880  
 R:Anzick, S.L.; Kononen, J.; Walker, R.L.; Azorsa, D.O.; Tanner, M.M.; Guan, X.Y.; Sautel  
 Science 277, 965-968, 1997  
 A:Title: AIB1, a steroid receptor coactivator amplified in breast and ovarian Cancer.  
 A:Reference number: 215053; M01D:97400625  
 A:Accession: T03749  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA

A:Residues: 1-1213,1218-1424 <ANZ>  
 A:Cross-references: EMBL:AF012108; NID:g2331249; PIDN:AA051677.1; PID:g2331250  
 R:Li, H.; Gomes, P.J.; Chen, J.D.  
 Proc. Natl. Acad. Sci. U.S.A. 94, 8479-8484, 1997  
 A:Title: RAC3, a steroid/nuclear receptor-associated coactivator that is related to S  
 A:Reference number: 214950; M01D:97385128  
 A:Accession: T03443  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-130,'EA',133-1213,1218-1273,1277-1424 <LIH>  
 A:Cross-references: EMBL:AF010227; NID:g2318005; PIDN:AA051663.1; PID:g2318006  
 A:Experimental source: cell line HeLa  
 A:Genetics:  
 A:Gene: TRAM-1; AIB1; RAC3  
 A:Map position: 20

Query Match 35.8%; Score 2729; DB 2; Length 1424;  
 Best Local Similarity 42.5%; Pred. No. 1,4e-127;  
 Matches 644; Conservative 243; Mismatches 451; Indels 176; Gaps 44;

Qy 1 MSGGENTSDPSRAETRRKRCPPQLGSPKRNTERKRNDEENKRIEELAFANENDI 60  
 Db 1 MSGGENTL-DEIADSDSRKRLPCDTPGGLTCSGSEKRRREDESKYIEELISANLSDI 59  
 Qy 61 DNFNPKDKCALKEVYQIRQIKOEKAAANIDVOKSPVSSGTGCVIDKDGPMML 120  
 Db 60 DNFNPKDKCALKEVYQIRQIKOEK- - - - -TISNDVDVOKAVSSGTGCVIDKDGPMML 118  
 Qy 121 EALDGFPEFVNLGNNVFESENVTQYLRYNOELMKNVSILHYGDHTEFVKMLPKSI 180  
 Db 119 QALDGFLEFVNRDGNIVSENVTVQYLQKQEDLVNVSIVILHEBDEKDFLKN-LRKST 177  
 Qy 181 VNGSGSEPPRRKSHFTNCMLYKPLPDESEBCHDNOEAHQKETMOCFVSPKSIKE 240  
 Db 178 VNGSVMTNETQKSHFTNCMLKTPDILIEDINASEPMQRYETMOCFALSOPRAME 237  
 Qy 241 EGEDLOSCLICVARRVPKRPVLPSSSEFTTRDLOGKITSLTPRMAKMGWEDLY 300  
 Db 238 EGEDLOSCLICVARRITGERTFPSPNPSFTTRDLOGKITSLTPRMAKMGWEDLY 297  
 Qy 301 RRCIOKFAHQEGESVSYAKRHHEVLRQGLAFSOIYRFLSDGTLYAAQTKSLIRSQ 360  
 Db 298 RRCIOKRFELSDGQSW-S-QKRHYQDVALNGHAELPYRFLSDAGTITYAQKSLFLNPV 356  
 Qy 361 TNEPOLYISLMLHREOVNVMNPDLTGQTMGKPLNPISSSPAHQALCSGNPQDMTSL 420  
 Db 357 TNDHRGEVSTHFLQREONGYRPNPNVQOQIRPDMACNSSVVG- - - - -MSMSPMQGLMP 412  
 Qy 421 SNINPPIGPKQEGMPGRGSGGGMNHS- - - - -GMQA-TTPQGSNYALKNNSPSQSPG 476  
 Db 413 SSRATGLADPSTTGOMSGARIGSSNIALSLTPGPGMOPSSYSTQNNNTGLNLSPPHSGPG 472  
 Qy 477 MNPQOPTSMLSPRHRMSPGVAGSPRIPSPQSPAGSLSPGVYCSSTGNSHSYNSSLNA 536  
 Db 473 LAPNOQNMISPRNR- - - - -GSPKIASHQSPVAGVHSPMASSGNGCN-HSEFSSSLISA 525  
 Qy 537 LQALSEGHVSLGSSLASPLDKMGNLONSPVNNMPPPLSKMSLSDSCDELYGE- - - - -PSE 594  
 Db 526 LQALISEGVTSLSTLSSPGPK- - - - -LDNTP-NMNIITOPSKVNSDSSKSLDFCYCDONPVE 581  
 Qy 595 GTTQOASSCHPGROKFNNDNPLPAYSSERAQOSRLHDSKGTGLQLQLTTTSSD- - - - - 650  
 Db 582 SSKQOANSRDLSD-KESKES- - - - -SVBEAENQGRPLESKGHKLLDQLTSSDGRH 634  
 Qy 651 -QMEPSPLASSL- - - - -SDTNKDSGTSLSGSSTHGTSLKEKHLILHRLDQSSS 699  
 Db 635 SSLNSPLDSSCKSSSVYSPSSVSSSTSGCVSTSMHOSLQEKHRLHLKLOGNS 694  
 Qy 700 PVDIAKTLAEATGKDLQOESSSTAPEGSEVTVKOEVPSPKKE-NALLRYLLDKDXTDIG 758  
 Db 695 PAEYAKITAEATGKDLQOESSSTAPEGSEVTVKOEVPSPKKE-NALLRYLLDKDXTDIDIG 751











[illegible]

```

Db      3600  -HLOOISLMSHG---qRKLtAqRPMGSLdQldQqQqQldQqQqQldQqQqQldQ 3655
OY      1216  VnLtlRPeVlPQAFiNAQMLAQOREtLl---NqHfQRQMHQOQOYQOFTLMMRGGLM 1277
Db      3654  QOL-----QOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 3705
OY      1273  TpsWAP--SGMPATMSNPRIQANAAQPFEPNNG-----ISQOPRG 1314
Db      3706  SRTLSLPQOQOQOQOVALGPGMPAKPILOHSSPGALGFTLLTGKEQNTYDPAVSSBATE 3766
OY      1315  -----FTGATTPQs-----PLMSBMAHTQSPM---M 1336
Db      3766  PstHOGGLAGlGTTPEsmATEPGEVPRSLdGSQSLdLVQRPQRPQSSlQldPRLPRLQ 3822
OY      1339  QOQOANPAIYQAPSDiNcMAQcNMGGNSMFQOQSPHPGQOANTSMISNNMNIYsMATPT 1399
Db      3826  QOQOQVSLTLHTA---GGSGHQOLdGSSSSBASSVPHLLAQP-----SVSLGDP 3877
OY      1399  GGMSSMNOQTQlSMtSVTSVTSGLSSMGPEQVNDPALRGGLPNDLPGMDMKQEQ 1457
Db      3871  GsM-TQNTLGPQPMLEPRMONTNGPQPPK---GVLQSG---QGLPGVIGIIPYNG 3920

```

```

RESULT      9
t13049
eyelid - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
A:Accession: t13049
R:Treisman, J.E.; Luk, A.; Rubin, G.M.; Heberlein, U.,
submitted to the EMBL Data Library, March 1998
A:Reference number: 217592
A:Accession: t13049
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-2715 <FRET>
A:Cross-references: EMBL:AF053091; NID:g2981220; PID:g2981221; PIDs:MAC06254..1
C:Genetics:
A:Gene: eld
A:Cross-references: Flybase:Fbgm0003013
C:Function:
A:Description: could act as a transcription factor antagonistic to the wg pathway
C:Keywords: DNA binding

Query Match          4.0%; Score 309; DB 2; Length 2715;
Best Local Similarity 20.7%; Pred. No. 4,6e-07;
Matches 259; Conservative 100; Mismatches 454; Indels 436; Gaps 56;

OY    392 GKPLNPIS-----SNSPAHQALCSGPPGQDMTFLSSNINFPIN--GPKDQMGRPRGG 443
           |||         |||         |||         |||         |||         |||
Db     675 GPPPPPQQAGAGAGCANSPSGAQAGGPGCGMHNHGTGPQWVWPSPPOOTVPQGARGG 734

OY    444 SGGNNHVSGMQATTPQGSNYALKMNSPQSRRP----GMNPQGPTSMLSPRHHSPGVAGS 499
           :|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db     735 AMVGNNHQGKCTPP-----PYVGGERPPRGSSSPRLAYLKLOHLHKKGYGGS 783

OY    500 PRIPSQSPAGSLHSFVGVCSTGNSHSTYNSSLNALQALSEGH-GVLSGLSLASLPDLK 558
           |||         |||         |||         |||         |||         |||
Db     784 P-TTPQ--GPGGYGNGCTGM-----HPGMFMG-----PPHH 811

OY    559 MGNIQNPSVNPNRP-----PLSKMGSLDKDCFLYGEP--SEGTTQAQESSCHPGEOKETN 613
           |||         |||         |||         |||         |||         |||
Db     812 MGP-PHGRTNNGPRTSTPPQSOH--LOG-----GDPQGQASGAPES---GPEHTS 857

OY    614 DPNLPPAVSSERRADGOGRLDSDSGQTQLQLLTTSKDSQMEPRPLASSLSDTNKDTGSLP 673
           |||         |||         |||         |||         |||         |||
Db     858 QDN---GISSSGPTGAAGMH-----AVTSVVT-----GPDGSMDEVYSQOSTLSNA 901

OY    674 GSGS-----THGTLSEKKHIIHLRLDDSSSPVDLAKTIAEATGKDLSQESSSTAP--- 724
           |||         |||         |||         |||         |||         |||
Db     902 SAAGGEDPQCTTPKSRKNDRYPQSGLHAPSTPHPVMMHGGGPEGVEYDMSSPNMPPRA 961

```

QY 725 GS-----EYTIQEP-----VSPKKENALLRLTLDKDQTKD----- 756  
 Db 725 GS-----EYTIQEP-----VSPKKENALLRLTLDKDQTKD----- 756  
 Db 962 GSPGVFNHVPVQEPFRSTITTTTKSDSLCK-LYEMDDNPDRRGWLKLRAMEERRPPI 1020  
 QY 757 ISLPETPR-----LRLDSKT----- 773  
 Db 1021 TACPTISKQPLDYRLIYVERGGFEYTKSKTKMDIAGLIGASSSAATLKRHYTK 1080  
 QY 774 -----DP-----ASNKLIAKTEKEMSEFPGDQPGSELDNLEIIL 811  
 Db 1081 NLTFECHPRDGRDIDPLPIIQVEAGSKK-----KAKAASVSPG-----GCHLDAGTNTST 1133  
 QY 812 DLQNSQ-----LPQLFPDTRPGAPAGS-----VDKQAIINDIMQTLAENSRYT 854  
 Db 1134 GSSNSODSPFPAPPGSAPRNALIDIGYPCGSPYPCGASGPDPYATAGQMQRPPSONNPQT 1193  
 QY 855 PVGAQKTLRLISQSTFNNRPRQGLRLLPNOMLPDITLOSPTGAPRPPIINSSPYSI 914  
 Db 1194 -----PHGAAAIVAAGDNISVSNPEPDLIAAGCGSTGCGPGOG 1235  
 QY 915 PGP-PMGNGMIGNOGNLGNSGTGMIGNSASRPTMPSGEMAPDSSAVRVTCATTSAMN 973  
 Db 1236 PGPGAASGAGAVGAVG-----GGQPHPRPH-----SPHT-----AAQQAAGQ 1275  
 QY 974 RPYVGCMIRNPAASTIPMRSSQPGQRQTLQSOVMNIGSELEMMNGGPGYSSQQAAPNQT 1033  
 Db 1276 HQOQHPQHHPRLPBPBPPOQOQGOOG--QDPSPSYG-----GCPBPAPQJHGPGQV 1325  
 QY 1034 APWPES-ILPTQASFAFASQNRQPFSSPDLLICHPRAEAPSPDEGALLDQLTLARNFDG 1092  
 Db 1326 PPSPOQHVRPAAGAY-----PPGSSGYPTPYSKTPGSEYRQPAIGQY-----G 1371  
 QY 1093 LEEDRALGIBELVYSQSAVDPEDPSODSNIM-----LEQ 1128  
 Db 1372 SSDQYNATGPR---GQPPGQGGQYPRQNRNMYPRYGEGEARPTGANGYGYGSRPSQ 1428  
 QY 1129 KAPVTPQ-----QYASQAMQAGSTSPMD-----PNF-- 1156  
 Db 1429 PPPGPPQPTQTVAGGPBPAGAPAPSSAYPTGRPSODDYQPPRDPDPSPPORRRHDFIK 1488  
 QY 1157 -----HTMGQRPSYAT-----LYMORPGLRPTGLVON 1184  
 Db 1489 DSGPRGTGNARQIYGANQSGTQYTRFQYPPSSPARQNMGGAPRRKRAAPRPG-APHGPRPQ 1547  
 QY 1185 QPNQLRLQLHRL-----QAQNRQPLANNQISN-----YSN 1215  
 Db 1548 QPAGVADMDQHNRYPRQGGPRPPRQOQOQPOQOQOQOQPRYQVAVGPPGQDPQAPQPMQAMN 1607  
 QY 1216 VNLTLRPV-PTQATINMQLAQRRELINQLRQRNHNQOQOQVOQRTLMHNGGLMTP 1274  
 Db 1608 PEGTARQSSILARPGSLARPPSGPQQRMRGMPARQOQOQOQGVPO-----PPR 1656  
 QY 1275 SMVAPSGPATMNSNRIPQAN-----AQQFPFPNNGISQOPRGTGATPOSPLMS 1327  
 Db 1657 QOASHGVV-----SPGLRQVGVGKWKPRYAPARPPSSQGVQGVQGGPPGGMMSKPRPM 1712  
 QY 1328 PRMAHTGSPRMAQOQA-----NPAYQAPSDT--NCGMAGNMGNSM 1366  
 Db 1713 PGOAMQOQPLQOQPRSHQNHPRHQHQRHQMPPNOTAPRGYVPRGM 1761

RESULT 10  
 H96747  
 unknown protein T10D10.14 [imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
 C:Accession: H96747  
 R:Thelodopsis, A.; Ecker, J.R.; Palm, C.J.; Federpiel, N.A.; Kaul, S.; White, O.; Alonso,  
 Chiu, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.,  
 ansen, N.F.; Hughes, B.; Hultzer, L.  
 Nature 408, 816-820, 2000

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziani, R.; Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A.Authors: Salberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A.Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.  
A.Reference number: A6141; MUID:21016719  
A.Accession: H96747  
A.Status: preliminary  
A.Molecule type: DNA  
A.Residues: 1-1088 <STO>  
A.Cross-references: GB:AE005173; NID:G6730762; PID:AAF27151.1; GSPDB:GN00141  
C.Genetics:  
A.Gene: T10D10.14  
A.Map position: 1

Query Match	3.9%;	Score 297.5;	DB 2;	Length 1088;
Best Local Similarity	21.3%;	Pred. No. 4.7e-07;		
Matches 259;	Conservative 146;	Mismatches 444;	Indels 365;	Gaps 65;

OY	377	ONCVAM-PLDT--GOTMKRPLNPISNSPAHAL-----CSCNPQODTLSSNINFPNG	429
Db	15	KVICIDRLPESSERGNLPGHLMQOTNNNOADIONLGTNNLADLRSQLODAPNSSLALVP	74
OY	430	PKEO--MGMPGRGSGSGGNHVS-----GMOATTPGGSNYVALKMNSPSGSGPMNG	480
Db	75	POOORIMIGSTRNTODGNSVSVSGASFGGIDALALPGSD-----SMPG	121
OY	481	QPTSMLSPRH-----RMS--GGAVASPRIPRPSQSPASGLSHSPVGCSSYGNSHSTYNSL	534
Db	122	-TSPHRRRESQEQGMSMGLMKRRRV--SHHGRCGVYQOOLGQ-RMDGLHGSDTNMKN	176
OY	535	NALDALSEGHCVSTGSSLASPLDKMG-----NLONSPVMMNPPLSKMSTL--D	581
Db	177	TLLQ-----HODMIGRSIQPNTSIOFRSPHMEGYMNOEGGPMOF---PASOGGMKYT	228
OY	582	SKDOFLGXPESEGTQOASCSH---PGOKRTEPNPLPPVSSERA-----	626
Db	229	SKE-----EPFE--TKIDGTRNNIPGVGSAND--LDPPIQSRMHPMAFINSNPOT	278
OY	627	-----DGOSRLHSGKGTKLLOLITTKSDME-----PSPILASLSDTNKDSGTG-----	670
Db	279	SWNVNPGQZLEKPKKEQGSRRISAO5PRLSAGRPQSPRLSKSGEPGSGSMGHNICAV	338
OY	671	-----SLPGSGTGTSLKEKHKLHRLDDSSPYDLAKLTAPATGKDLDSQSS	720
Db	339	AAOKRAVYISIPAGIQTOGVSSANEAMQORHQ-----AQMAKRRTNSLPKTYO	390
OY	721	STAGSEVTK--QEPSPKKENALLRLLDKDKDKIDGLPEIRPK-----LERDS	771
Db	331	ISTYSGPVSWNTISVYNAMS-----PSVGPOTLGDHALIDRESK	430
OY	772	KTDPA5TKLIAMKTEKEMSFER---GDOPGE-LDNL--EELIDD-----	812
Db	431	IERVAARQOLNCKKHKXDEYSRRPRYAKOPLTCLSLNSNEVEPKDEDEALSKSIFGGS	490
OY	813	-----LONSQJ-----POLPPDR-----PAAPA---GSVDKQ	837
Db	491	MNTYKTRVHFGOMERYMOSSIFSGSVSFLPRNTRTLWSEKAVDGTVMYOGDVE--	549
OY	838	AIINDLMQTLAENSPLY--PYGAOKTALRISQSTFNNPRPGQGRLLPNOJNPLDTITQSP	896
Db	550	---GDVQ--AEDFLALPPTMARREGYMLDEHIMAKPRNGDYG-----PIS---SHP	593
OY	897	TGAGPFP--PIRNSPYVYIPQPGMGNMGIMOGNIGNSST--GMIGNSASRPTMPSG	952
Db	594	NSAGGYPRGYSANDMOQYGAVALVGAQSGASKHGNTGNTPNNSTONILLANRMP--PTN	651
OY	953	EMAOSSA5AVYTCALITTSAMNRPVYOGGIMRNPRA5TIPARPSQPORQORLOSQYMNITGFS	1012
Db	652	SQALOMSO-----GLL--SGVSPMPQOOLDOJQ5ALLS-----S	684
OY	1013	ELENMGGP00SO00APPNOTAPWPE5IIP0ASPASONR0PFPSSPDILG-----	1065



```

Db 2244 --000000000000GWPSPHPTPSSKSSQASVPSQTAMAAHQBN-----YQAKPPQMSA 2295
Qy 1190 RLQLOHRLQAOONRPLMNIQSNVSNVLTLPVY-----TQAPINAOHLAORQ 1239
Db 2296 MS000HSMQATPTQOPHA-----LGLREPAIRGQAVFSAHEAOSPISVSHOH 2345
Qy 1240 REI-----LMOHLRORQ----- 1251
Db 2346 KSLGRSGSPFPPMRDPRQONIRKGEFVPPOGOPYRYRYNTGPRGCGGPGGPGVPGCE 2405
Qy 1252 -----MHQOQOYOQRTLMKGQGL-NMTPSMVAAPSGPATMSNFRIPA----- 1294
Db 2406 APADLRMQMSQARSYTPGPGAGFEGGPPSSSLGYPEQIRQADQIREMSRDL 2465
Qy 1295 -----NAOQFPPPPYIGI-SQOPDDEGTGAT-TPOSP 1324
Db 2466 GRDRHYMGDRPRELGRPREVAAQOREQHAQFOAQOQHGVAPANAHPQHIQVQGHPOH 2525
Qy 1325 LMSPRMAHTQSPMMSQOANPAYOAPSDINGMAQGNMGNSWFSQGSPPH-FGQOANTSM 1383
Db 2526 QIAHQHQHVQOQOQHDMRN-AVQA-SHQOYVPQHOGG-MAARQLRQHOTOYDOUGHGPG 2582
Qy 1384 YSNMNNIN-VSMAT-----NTGMSMSMQMTQO 1410
Db 2583 PANYMGADTIALTLRKNGTDGIFTSNKRHGO 2614

```

## RESULT 12

```

T30160
hypothetical protein C37A2.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T30160
R:Le, T.T.; Kemp, K.; Scheet, P.
submitted to the EMBL Data Library, April 1997
A:Description: The sequence of C. elegans cosmid C37A2.
A:Reference number: 220746
A:Accession: T30160
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1234 <LET>
A:Cross-references: EMBL:U97194; PIDN:AB52447.1; GSPDB:GN00019; CESP:C37A2.2
A:Experimental source: strain Bristol N2; clone C37A2
C:Genetics:
A:Gene: CESP:C37A2.2
A:Map position: 1
A:Introns: 46/1; 124/1; 159/1; 231/1; 508/3; 740/3; 891/1; 931/1; 975/1; 1014/3; 1083/3

```

```

Query Match 3.88; Score 287; DB 2; Length 1234;
Best Local Similarity 21.08; Pred. No. 1.9e-06;
Matches 222; Conservative 111; Mismatches 366; Indels 336; Gaps 54;

```

```

Qy 572 PPLSKMGSLL-DSKDCFGLYGEPSEGTG-QAESSCHPEQKETNDPNLP-----AVSSE 624
Db 17 PDIEKVGDIYVPSATALLSLSTSSSGIEASTGC---ASTPTANTLPPLIEESVDE 73
Qy 625 RADGOSRLHDSKQTKLQILTTSDQNEPSPLASLSDTYKNDSTGSLPGSGSTHSTSLK 684
Db 74 TADVGIAIE-----KLV-----KSLIVPSP-----SDTAR-VTSAISNQTVAASKTNE 117
Qy 685 EKHILRLLODSSSPVDLAKLTAEANG-KDLSQESS-----TAPSE-VTIKQEVSPKK 739
Db 118 EDRKSI-----DTSVDLQEREEDDTGDNOISQOSSSKDHECEPREDLSTHEEPTNSPT 172
Qy 740 KENALRLYLDKDDTKDIGLP-----ETPKLE-----RLDSKTDPSNFKLIAMK 785
Db 173 NDGYL-----PSTSGIRASPADMTASSRNPLRKRYSLRLPAKA--AQRLKRIR-- 220
Qy 786 TEKEBMSPEPD--QPG-----SELNDLEILDDLQNSQLPQLFPDTPGAPACSVDKQ 837
Db 221 -----NFQPSLRIRIGLSALKISDMCLR--LADGSPQIPVLEED-HINIPSTSQENP 271

```

```

Qy 838 AIINDMLQTAENSPVTPVGAOKTALRISQSTFNNPR-----GOLGRL-----PMQN--- 886
Db 272 RDIKSLSHOLPMIPPOQPPPPVPSHMLPSSTGPGHPSHMGRLSQDLLPSGPMQGHNS 331
Qy 887 ----- 886
Db 332 PQVMVKQEPESQFTPQPHMQOTPOQOYLQYPRPMQPHQNMQNMQNTAEYQAQMRAG 391
Qy 887 -LPDLITLQSPGTAG-PPRPLINSSPYSVIDPQGMNGMGITGNQNLGNSSTGMIGNSA 944
Db 392 FMAAQIKQEVSGSGQPPPVGEPQPOQITPQGSILGPMGSLG----- 434
Qy 945 SRRTPMSE--WAPSSAVRYTCATTSAMRNRYGGMIRN--PAASTPMRPS----- 993
Db 435 -PTAPPGSQPMNPQOQRKIQOQOQAAPSASNSPLLVNLSNQOQPOQOYMTGPSAQGLS 493
Qy 994 -----SQPGQRTQLQSOVNIGPSELEMMNGGPOYSQOQAPRNPQTAP--WPESILPI 1043
Db 494 MQQIALIQOQOQHQOYLQRIQQOQQAAMQOQOQOQOQOQOQAPRPTPRNGHPOGFPT 553
Qy 1044 DQASFASQNRQPPGSSPDDLCPHRAESPS-DEGALLDQYLALRNFDELEIDRALGI 1102
Db 554 NQAG-----PPG-TPGRPIPYAMGQPRMYHQAGPOGOMIQRMNSYPRGNAQDPR--- 601
Qy 1103 PELVQSQAQVDEQFSSQDSNIMLEQKAVFPFQYASQAQMAQSSVSPMODPNTFMQR 1162
Db 602 PPGQSQQPIPPQO-----QOAPRAPPQOL-----QOAPVQDT----- 634
Qy 1163 PSYATLRMQPRPGLRPT-----GLVQNPQNLRLQO---HRLQAOONRQPLAMQI 1210
Db 635 ---SAVAERPKKKRPTKKQKEAALALEQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 690
Qy 1211 SNVSNVNLTLRPGVPTQ-----APINQMLAQORETLNQLROR 1250
Db 691 -QOQOQOQVQOQPGYQOQYPCGMPRPGQAFPRQYELGGAQPTQOQIQOQOQOQOQOQOQO 746
Qy 1251 OMHQOQOQOQRTLMRSGGLNMTPSMVAAPSGPATMSNFRIPAQANAQFPPRPNYGISQO 1310
Db 747 AMFQOQOQOQO-----GQ-----OPGGQOQPMG--QWPOQOLRPVYPR--GNNST 786
Qy 1311 PDPGFTGATTQSPPLMSPRMAHTQS-----PMQOQOQANPAYOAPSDI---NGMAQGNMG 1362
Db 787 PD-----AVNQOQNPPIPGATMQHRLSGEFARPPVSSGKYNPQOQOQGSFHRSDSSASVYSG 842
Qy 1363 GNSMFSQGSPPHF-----GQOANTSMYSNMNNINVSMAITNGMSSMNO--MTGQISM 1413
Db 843 SHTPFGQOQGPQSQSEPTAVPQPOQN-----NPGGSGDIGEKAIVDQLLN 886
Qy 1414 TSVTSVSTSGLSMG-----PEQVND---PALRG 1440
Db 887 SSEPLADIGLDLDDLDIEPMQVODGQTPSTSG 921

```

## RESULT 13

```

T22330
hypothetical protein F47A4.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T22330
R:Mortimore, B.
submitted to the EMBL Data Library, June 1995
A:Reference number: 219549
A:Accession: T22330
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3498 <NLL>
A:Cross-references: EMBL:Z49888; PIDN:CAA90064.1; GSPDB:GN00028; CESP:F47A4.2
A:Experimental source: clone F47A4
C:Genetics:
A:Gene: CESP:F47A4.2
A:Map position: X
A:Introns: 36/3; 72/3; 147/2; 228/3; 261/2; 320/2; 847/3; 881/1; 1317/2; 1500/3; 1653

```





```

Db 332 -----OFGNDLSGVDPGSGGYGOMNNSNTPNQSERSTPCQFSTP----- 371
Oy 751 KDDTKDILPETPKLELDSTKDPASNTKLIAMTEREKMEFEGDDPGSLDMLLEIL 810
Db 372 -----GTP-----GSHGTGVSNAPLSHOK-----PSQOQSMHNL 402
Oy 811 DDLQNSQLPOLFPDTPRPGAPAGSVDKQ-----AIIINDLMQLENSPVTPVGAOKTALRIS 866
Db 403 PHHOVNQNNLLSPN--HASSLSGOKOHGSSPMSSILMPINQYQSMKQ--NNQASASISM 459
Oy 867 QSTENNPFGQLGRLLPNQNL--LDITLQSPFGAG--FPPIRNSSPYVTPQDGMGN 922
Db 460 EPTFEKP-----AVPIRHSFSPQMTLQSVHSPSPGAPAPAYANAPSSSKTPDP----- 507
Oy 923 QGMIGNQGNLNSSTGMIGNSASRTMPSGEMAPQS--SAVRYTCATTSAMRPYQGMKI 981
Db 508 -----TQOQRHSPTEFAV-----PTLPAATLQAQAFSANQISTPKTSPQKKHEDGVP 556
Oy 982 RNPAASIPMR-----PSSQPGOROTLOSQVMNIGPSELEMMNGGPQY-----SQOQAPPN 1031
Db 557 EPPRADPFTVTYTHLELPAATFLRDTL-----HYGPDKXHPQVEKHYSFKRQQL--- 608
Oy 1032 QTAWPESI-----LPIDQASFASQNRQPFSSPDDLCPHPAESPDEGALLDQLYLA 1086
Db 609 -RVYRPEGINSHTTPTEPNTFEGFMQGNPYDPKYNRVVP-----SQTSHG----- 653
Oy 1087 LRNFDELIDRALGIPRELVYQSOAVNDPEQSSQSDSNIMLEQKAPVFPQYASQQAQAG 1146
Db 654 -----PLLSRSQSM----- 663
Oy 1147 SYSPMODENFH-----TWGQRPVATLRMQPRGLRPTGLVONOPNQLRLQHLQAOQ 1201
Db 664 -HTMISPFNFASQSPSTSGRQF--AKKARSASDASEPPFPNPHRPSGSGMDQRLQOO-- 719
Oy 1202 NRQPLANOISVSNVNLTLRPGVPTQADINQMIAQQRREILNHLRQRMHQOQ--VQO 1260
Db 720 -----QIOMQOYHQHMQKMQOQOQMAAQ 744
Oy 1261 RTLMNRGGGLMTPSPMVAPSGMPATMSNPRIQANAAQFPPRPNTGISQOQPRGTGATY 1320
Db 745 QMSRMKGGSG-----PSSAGPQG-----SOLPS----- 766
Oy 1321 POSPLMSFMAHTQSPMMAQSOANPAYQAPSDINGMAQNGMNSMFQSQSPPHFGQOAN 1380
Db 767 -----LSAPSLQRADSNP--QLP-----SQOQPRMGGRMAN 795
Oy 1381 TSMTSNNMNINVSMTATNGMSSNMQMTGQ--ISMISVTSVTSIG-----LSSMGPE-- 1430
Db 796 -----HMGGMQPMNGPTPEGPTVNNIGLNSNMAAGLPLSLRSQGPDS 839
Oy 1431 QVND 1434
Db 840 QNND 843

```

RESULT 15  
A54277  
transcription adaptor protein p300 - human  
C:Species: Homo sapiens (man)  
C>Date: 09-Sep-1994 #sequence\_revision 09-Sep-1994 #text\_change 10-Dec-1999  
C:Accession: A54277; S60344  
R:Reckner, R.; Ewen, M.E.; Newsome, D.; Gerdes, M.; Decaprio, J.A.; Lawrence, J.B.; Livingston, D.; Ewen, M.E.; 1994  
A:Title: Molecular cloning and functional analysis of the adenovirus E1A-associated 300-kDa protein p300  
A:Reference number: A54277; MUID:95011587  
A:Accession: A54277  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-2414 <ECK>  
A:Cross-references: GB:001877; NID:g495300; PIDN:AAA18639.1; PID:g495301  
A>Note: In the authors' translation 941-Ser is shown after 961 and consequently, residue R:Lundblad, J.R.; Kwok, R.P.S.; Laurence, M.E.; Hartner, M.L.; Goodman, R.H.

```

Nature 374, 85-88, 1995
A:Title: Adenoviral E1A-associated protein p300 as a functional homologue of the tran
A:Reference number: S60344; MUID:95174889
A:Accession: S60344
A>Status: preliminary
A:Molecule type: protein
A:Residues: 552-660 <LUN>
C:Genetics:
A:Gene: GDB:EP300
A:Cross-references: GDB:9862958; OMIM:502700
A:Map position: 22q13.2-22q13.2
C:Superfamily: unassigned bromodomain proteins; bromodomain homology
C:Keywords: phosphoprotein; transcription; zinc finger
F:1075-1132/Domain: bromodomain homology <BR>
F:89, 507, 1136, 1295, 1497, 1834, 1977, 2062, 2320/Binding site: phosphate (Ser) (covalent)
F:1734/Binding site: phosphate (Ser) (covalent) (by protein kinase A) #status predict

```

Query Match 3.6%; Score 277; DB 2; Length 2414;  
Best local similarity 18.9%; Pred. No. 1.5e-05;  
Matches 222; Conservative 117; Mismatches 363; Indels 474; Gaps 50;

```

Oy 361 VNPDLTGOTMGKPLNPI-----SSNSPAHQALCSGNPGQDMLSSNINFPING 429
Db 47 IINSTELGLTNGDINOLOTSIGVYQDAASKHQLSILLRSGS-----SPNLNMGYGG 99
Oy 430 PREQMGMPGRFGSGSGGNHVSQMATTPOGSNTYALKNNSPQSSPGM-----NPGQ 481
Db 100 PQQVNA-----SQAQOSSPFGGLINSNVKSPMT 127
Oy 462 PTSMLSPRHRMSPGVAGSPRIIPSOEFPAGSLHSPV---GVCSTNGNSHSYTNSSLN-AL 537
Db 128 QAGLTSP--NMGMGTSG--PNOGPTQ--STGMANSFVYNPAMQMTG-----TNAQMGNM 177
Oy 538 QALSEHGCV-----SLGSSLASPDLK-----MGNLQNSFVNNNPPLS----- 575
Db 178 IAAAGGQGIIMPQVANGSIGAGRGQDMQYPNPQMGASGAILTEPLQOGSPQMGQGTGLR 237
Oy 576 -----KMSLSKDCFGILXGEPSECTGOAESCHPQKQEN---DNLPPRAVSSERA- 626
Db 238 GPQPLKMGMMNPN---PYGSPYTONPGQOIGASGLQIDQTKYLSNNLSFPAADKRAV 294
Oy 627 --DQSRLLHDSKQTKQLLITTKSDQMEP--SPLASLSLTDNKTSGSLPGSSGTGT 681
Db 295 PGGGMNPN---GQQPARQV---QQPLGYTPVAQ-----GKSGAHT 329
Oy 682 SLKEKHKLHRL-----QDSSSPVD-----LAKLTAATGKD- 714
Db 330 ADPEKRKLIOOQLVLLHAHKCORREQANGEVQCNLPHCRTMKVNLNHTHCQSGSCQ 389
Oy 715 -----LSQESSSTAPGSEVYIKQEPVSPKKENALLRYL---LDKDTKIGLPE 761
Db 390 VAHCASSROIISHMKNCYTRHDCPVCLPLKNAGDKRNOQPIITLGAPVGLGNPSSLGVG-QQ 448
Oy 762 ITPKLERLSDKTDPAASNTKLI-----MKTE-----KEEMSFEPDQPGS--ELD 804
Db 449 SAPNLSTY--SQIDPSSIERAYAAALGPYQVQNMPTQPOVQAKNQOQNGQPGQSPQGMKMS 507
Oy 805 NLEETLIDLQNS--QLPOLPDTTPRPGAPAGSVDKQAIINDLMQLENSPVTPVGAOKT 861
Db 508 NMSASPMGVNGVGVQTPSLTSDSM-----LHSAIINSONPMSASVPSLGPMPPT 558
Oy 862 AL-----RISQSTFNNPFGQLG----- 879
Db 559 AAQPTGTGIRKQWHEDTIDQLRNHLVHKLVAQIFPTPPPAALKDRMENLVAYARKVECD 618
Oy 880 -----RLLPNQLPDLITLQ-----PTGAPFPPIRNSPYVTP 915
Db 619 MYESANNRAEYTHLLAEKTYIKQKLEKRRRLQOKOMLPRAGAGMYVSNAPGPMQOP 678
Oy 916 QPMGMGNQMTIGNQGNLNSSTGMTIGNSASRTPMPSGEMAPQSSAVRYTCATTSAMNRP 975
Db 679 QPGMTSN-----GLPDPDS--MIRGSVPPQNMPP----- 704

```



Db 2068 SMMPPEARS-----LYEOPGGLSSYINRDSIGATNGVLTLPQOS-IQRPSTASTSSN 2119  
 QY 1124 IMLEOKAPVFPQOYASOAOAGSYSPMODNPHFMGORSYATLRMQPRGLRPTGLVQ 1183  
 Db 2120 ---PPKAPLDP-----SASVONNTIDPAIEIEIRV--QRMFTKPLKMSAEAAIYMAVAS 2169  
 QY 1184 NQPN-----QLRLQLOHRLQAOQNRQ-----PLMN--OISNVSNVL 1218  
 Db 2170 SDPMPRATSTVDLAAMQ-QQAAQAAQAAQOQVPTTASTPRLSNLETLLSTASLANL 2228  
 QY 1219 -----TLRPVPTQAPITNAQMLAORRELLNQLRQR--OMHQOQOVQ 1259  
 Db 2229 ATGALNPLSMALTSSTLQSSPYOGIARVLLTMNMCOMLATQTSSELLATMNOQETLM 2288  
 QY 1260 QRTLMNGOGL-----NMPSPVAVPSC--NPATMSNRLPQANAQOF----- 1299  
 Db 2289 ---ALLAARNGLPPFAMPQOQOQOPMPAOGGFAIPVLPRLMSIKRNAKDDLSTGVGVSDRKS 2346  
 QY 1300 -PPPPNYGISQOPDPGFTGATTPQSPPL--MSPRAHRTQSPMAQOQOANPAYQAPSDINGW 1356  
 Db 2347 CPLHAMIQGGQOPPP-----POQPMQAVAPARPPSPPRKSMFEN--LPEPMK-- 2393  
 QY 1357 AQGMGNSMF-----SQOSPHEGOQANTSMYNNMNIIVSMAT 1396  
 Db 2394 -----EKEMERKEILRLDILILELGAEDDEQKPDLKQ-----IPTSEDTDDSKAD 2443  
 QY 1397 NTGG-----MSMNMTGOISMTSVTSYTSGLSMSGPQVNDPRLRGNLFPN 1445  
 Db 2444 SMGEGSAFRLIRLSRSTSMGNSGSPASGTTSPSTSSISSPD--SPLEGE--PL 2497  
 QY 1446 QLPQMDMIKQEGDTTRK 1462  
 Db 2498 NSEFMDLTFEVAQKHRE 2514  
 RESULT 17  
 B86292  
 F7H2.12 protein - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 02-Nov-2001  
 C:Accession: B86292  
 R:Theologian, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
 ansen, N.F.; Hughes, B.; Hultzer, L.  
 Nature 408, 816-820, 2000  
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,  
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
 A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A:Reference number: A86141; MUID:21016719  
 A:Accession: B86292  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1366 STOS  
 A:Cross-references: GB:AE005172; NID:98927657; PIDN:AAF82148.1; GSPDB:GN00141  
 C:Genetics:  
 A:Map position: 1

Query Match 3.64; Score 273.5; DB 2; Length 1366;  
 Best Local Similarity 18.54; Pred. No. 1e-05;  
 Matches 252; Conservative 186; Mismatches 459; Indels 461; Gaps 63;

QY 261 RVPVPSSESTTRDLOGKITSLTSTMRAMKPGMEDVLRICQKFAHQEGESVSAYAK 320  
 Db 7 RPLSPNCE-----PAMDTGDMRQLRP--DSRQIKYVKI----- 38  
 QY 321 RHHEVYLRQGLAFS-----QYRFSLS-DGLVAAQTKSLINSQTTN----- 362  
 Db 39 ---METLKHLFPFGREGINELRIARAFEEKIFSGALNQRFQWMTPOHGKEITFGICK 95

QY 363 -EPOLV-----ISLHLHRE--QNVCVNPDLTGOTMGRPLNPISNSPAHQ 407  
 Db 96 AKPOVYGEIHFDYLRKLSMKLMTETKSQAAGSSAIPRANNTSTDSITPND--COL 153  
 QY 408 LCGNPGQDMLSSNINFPINGPKQMGMPGREGSGGMNHWGQMTTPQGSNYALKM 467  
 Db 154 L---PG---SLSTMQS--QAPQLSOTWON-----MTASQMTGST-----ALPS 190  
 QY 468 NSPQSSSGMNPQOPTSLSPRHRSFVAVASPRIPQSFSPAGSLHFPVVCSTGSH 527  
 Db 191 SMPVYS--TINNNTSVNON-----AMQVVAQMLDSSGQ 227  
 QY 528 SYTNSLIA--LQALSEGHVS-----LGSLSPLDKMGNLQNSPVNMP 572  
 Db 228 GLSSMFGPQROMLGRHAMSQQOQOPRYLXQDLQOOLKONFOSGNVNR-FNLSLPS 286  
 QY 573 PLSKGSLSDKDCGLYGPSEGTGQAESCHPEQKRETNPNLPPAVSERADGOSRL 632  
 Db 287 HIOQ-----QQQNVLDQPNQLHSSQOPGVPS----- 312  
 QY 633 HDSKQOTLLDQLTTKSKMERSPLASLSPTNKDSTGLSGSGTHGTSLKEKHILHR 692  
 Db 313 -----ATQPSVNSAPLOG--LHTNDOSSPOLSSQOYTQOS-----M 346  
 QY 693 LLDSSSPVDLAKLTAEATGKDLSESSSTAPGSEVITKQBPVSP-KKENALLRYLLDX 751  
 Db 347 LRQHOSSMLRHPQSOQSG--HQOQSS-----LPQOSIFLQOQPPOLNR----- 391  
 QY 752 DQTDIGLPEITPKLERLSKTPDASNTKLAMKTEKEMSEFERODPGSELDLEETLD 811  
 Db 392 -----QQANSSGI--QKQMG-----QHYVG 412  
 QY 812 DLQNSQLPQLPDRPRCAPACSVDKQAIINDLMQLTAENSV-TRVCAQKALRISOSTF 870  
 Db 413 DMQOQHOQL-----LQONNVNMIQOQOQOQOPLQOPOQOQK----- 450  
 QY 871 NNPRGQLRLPNQNLPLDTLQSPGAGFPPIRNSPSPSVIPOPQMGNOGIGNOG 930  
 Db 451 --QOPPAQOQLMSQON-SLQATHQPLG-----TQSNVAGLQOP--QOQMLNSQ- 494  
 QY 931 NLGNSSTMGNSASRPTMGEMAPQSSAVRYTCATTSAMNRPVQGMIRNPAISPM 990  
 Db 495 -VGNSSLQNNQSHVMLSQ-----TVGLQTRHQAG--HGLYSSQGG 533  
 QY 991 RPSSQPGQQR--TLQSYVNIIG--PSELEMMKGCQYSGOQAPPPQGTAPWESILP- 1042  
 Db 534 QSONQPSQOQMMRPLQSHHQDLGLQOQPNLLQOQV--QORLQASQGVGT--SLLRP 585  
 QY 1043 ---IDQASFASQNRQPFSSPDLLCPRHAESPDSDEGALLDOLYLARN-----PQGLE 1094  
 Db 586 QNVVDQQRQLYQOQRTLEPMPSSSLDSTAQTES-ANGQDQOEYVQKTKSMKETYLPLDN 644  
 QY 1095 EIDRAL-----GIPELVSOQAQVDEPQF-----SSQDSNIM-LEOKAPVF- 1133  
 Db 645 EITQRYVAKLQODSDMPQQRSDQLEKLRQFKTILERMIOFISVYSKSNIMPRALKQVAYTE 704  
 QY 1134 -----PQOYASQAOQAQGSYSFMDPNFHTMGQR-----PSYATLRMQ----- 1171  
 Db 705 KQIIGFLNMHRPRKRVQOQGLRPOSQOMQORPOSQOTVQDOSHDQNTQPNQMSKMGAGP 764  
 QY 1172 -----PRPG-----LRRTGLVQONPQNDLRQLQDQHL-QAQONR 1203  
 Db 765 RAQOSSMTNMOSNVLSRPGVSAPQONTPSSIPASSLESQGNLNNQOAVKMSMOQNT 824  
 QY 1204 QPLNQISNVSNVNLTLRPGVPTQAPITNAQMLAO-----ROREIILNQH 1246  
 Db 825 SQLVNNSSASASQSGISTLQSNVNVQPOLSSSLQNHQHLKQOQDQOQLKQOQOQOMQOQ 884  
 QY 1247 LRQOMHQOQOQOQOTLLMRQOGLAMTPRSVAVRSGMRAT--MSPRIRIQAANAQOPPPPP 1304  
 Db 885 LQARQOQOQOQOQOLARQOAAQOLQOMNDMLDTSRQGMNVSRRMFQOHSQOGARVPL--- 941  
 QY 1305 YGISQOPDPGPGTGAFTTPQSLMSPRMAHRTQSPRMMQOQANRAYQARSDINQWAGNGGN 1364







A:Cross-references: EMBL:U22431; NID:g881345; PIDN:AAC5152.1; PID:g881346  
 A:Note: parts of this sequence were confirmed by peptide sequencing  
 R:Hogenesch, J.B.; Chan, W.K.; Carver, L.A.; Bradford, C.A.  
 submitted to the EMBL Data Library, June 1995  
 A:Reference number: H00692  
 A:Accession: G01875  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-826 <HOG>  
 A:Cross-references: EMBL:U29165; NID:g1144012; PIDN:AAC51210.1; PID:g1144013  
 C:Genetics:  
 A:Gene: GDB:HIF1A  
 A:Cross-references: GDB:512229  
 A:Map position: 14q21-14q24  
 C:Keywords: heterodimer

Query Match 3.5%; Score 269.5; DB 2; Length 826;  
 Best Local Similarity 19.3%; Pred. No. 7.9e-06;  
 Matches 186; Conservative 150; Mismatches 337; Indels 289; Gaps 41;

OY 1 MSGKEINTSDPRATRRKKECPDQIGSPKNTKRNREONKYTELEAEILFANFNDI 60  
 DB 1 MEGAG-GANDKKRISSEKKE-----KSHDAARSRKSESEVEYELAHQLPLPHNVS 51  
 OY 61 DNFNFKPPKCALIKETVAKIROIKEOEKAAANIDEVOKSDVSTGGQVIDKDALGPMWL 120  
 DB 52 SHL-----DKASYMRLITSLR-VKRLDAGDUDIEDMKAKQNNC-----FYL 93  
 OY 121 EALDGFVVNLGNVFNVESENVTQYLRYNDELMNKSYSYLHVGDTHEFYKNLLPFSI 180  
 DB 94 KALDFVAVLFDDGDMIYIDSNVKNKMGTLQELFGHSVDFETHCDHEMKEMLTNR-- 151  
 OY 181 VNGGWSSEPPRRNRSHNTNCRLVYPLRDSSEGHNDQEAHKYETMOCFANVSQKSIKE 240  
 DB 152 -NG-----LVK-----KGEQNTQNSFFLRKSTLTLSRRTNMI 184  
 OY 241 EG-----EDLQSLICAVRVPKKEKRVLP-SESEFTT 272  
 DB 185 KSATKVLHCHGHINHYDTNSNQPGYKRRPMTCVLICERIPRPSNETIELDSKTFIS 244  
 OY 273 RODLOGKITSLDTS-TMRAAMKPGMEDIVRRCIOFHAOEGESVYAKRHHNEVLROGL 331  
 DB 245 RHSLDMKFSYCDERITELMGYER--EELLGRSIEYEU--HALDS-DHLTKTHDMFTGQ 299  
 OY 332 AFSQIYRRSLSDGTLYAOTKSKLRSQTNPOLVISH-----MLAEQNV 380  
 DB 300 VTTGQYRLAKRGYVWVETQATVLYTNKNSQPOCIVCNVYVSGIIOHDLIFSLOQTEC 359  
 OY 381 VMNP-DLTGQTFMKPLNPLISSNSPAH-----QALCSGNPGQDMTSSNINPPI 430  
 DB 360 VLKPYESSDMKMTQLETFYESHDTSSLFDKLKEPDALTLAPAGDTIIS-IDF----- 413  
 OY 431 KEQMKPMGRFGSGGMNHVSGMAQTTPOGSNYALKMNSPSSPGMNPQOPTMLSPRH 490  
 DB 414 -----GSDNETEDDOQLFEVPLVN---DVMLPSNE 441  
 OY 491 RM-SPGVAGSPRIIPSPQSPAGSLSPYGVCSSTGNSHTSSNALQALS-BEHGYSL 548  
 DB 442 KLQNTINLAMP-----LPTAETPKPL-----RSSADPALNOEVALKLENPESL 485  
 OY 549 GSSLASPLDKMGNLQNSFVNMPPLSKMGSLSDSKDGLGVEPSEGTGGAESSCHGE 608  
 DB 486 ELSTFMQIDQ-----DQTP-----SPDGSIRQ--SSPEPNS 515  
 OY 609 QKE-----TNDPNLPRAVSSERADQGSRLHDSKQTKL-IQLLT-----T 647  
 DB 516 PSEYCFYVDSDMVNEFKLELVEKLEAEDTEAKNPSTQDITDLMLAPYIPMDDFOLR 575  
 OY 648 KSDQMEPSPLASSLSDTKDSTGSL-----PSSGSHGSLKEKHILH----- 691  
 DB 576 SFDQL--SPLESSSASPEASPSQSTVTVFOQTQIQEPTANATTTATYDELKTVKDRME 633

OY 692 --RLDSSSPVDLAKITPAEAT-----GRDLQSESSSTAGS----- 726  
 DB 634 DIKILASPSPHIHKEITTSATSSPYROTQSTASPNAGKVIQTEKSHRPSRNVLSV 693  
 OY 727 ----EVTIKQEPVSPK--KKNALLRYLLDKDDT--KDQIGLEIPPKLERLDSKTDPSAN 778  
 DB 694 ALSQRTTYDEELNPKIALQNAQKRKMEHSDSLFOAVGIOTL-----LQOPDHAT 747  
 OY 779 TKLIAMKTEKEMSEFPGDQPSSELDNEELIDDL-----ONSOLPOLFP-DTRPG 828  
 DB 748 TSL----SWKRVKGCSSBQNGMEQKTIILLPSDLACRLLAGSMDSEGLPOLTSYDCEVN 803  
 OY 829 AP 830  
 DB 804 AP 805

RESULT 23  
 JC5405  
 brain and muscle Ah receptor nuclear translocator-like protein, BMAL1b - human

C:Species: Homo sapiens (man)  
 C:Date: 10-Jun-1997 #sequence\_revision 18-Jul-1997 #text\_change 01-Dec-2000  
 C:Accession: JC5405; JC5404  
 R:Ikeda, M.; Nomura, M.  
 Biochem. Biophys. Res. Commun. 233, 258-264, 1997

A:Title: cDNA cloning and tissue-specific expression of a novel basic helix-loop-helix site usage.  
 A:Reference number: JC5404; MUID:97289529  
 A:Accession: JC5405  
 A:Status: nucleic acid sequence not shown  
 A:Molecule type: mRNA  
 A:Residues: 1-626 <IKR>

A:Cross-references: DBJ:AB000812; NID:g2094734; PIDN:BAI19935.1; PID:g2094735  
 A:Accession: JC5404  
 A:Status: nucleic acid sequence not shown  
 A:Molecule type: mRNA  
 A:Residues: 'MINI', 48-415, 'P', 417-626 <IKR>  
 A:Cross-references: DBJ:D89722  
 C:Comment: This protein plays a role in the control of central nervous system develop

C:Genetics:  
 A:Introns: 47/2; 223/1; 250/2; 274/1; 288/2; 361/1; 392/1; 442/2

Query Match 3.5%; Score 269; DB 2; Length 626;  
 Best Local Similarity 24.3%; Pred. No. 5.6e-06;  
 Matches 142; Conservative 82; Mismatches 230; Indels 130; Gaps 27;

OY 35 EKRNEQENKYTELEAEILFANFNDIDNFNFKPCALIKETVAKIROIK-EQEKAAAN 93  
 DB 81 EKRNRDKMNSFLDELASLV---PTCNMMSRKLDKLYLRMAVQHMRLRGATNPTYEAN 136  
 OY 94 IDEVOKSDVSTGGQVIDKDALGPMLEALDGFVVNLE-GNVVFNSENVTQYLRYNQOE 152  
 DB 137 YKPTFLSD-----DELKHLILRAADGFLVVGCDRGLILFVSESVFKILNYSQN 185  
 OY 153 ELMNSVYSYLHVGDTHEFYKNLLPKS-----LVNGSGWSGGEPPR 192  
 DB 186 DLIGOSLFDYHPKDIAR-VKEQLSSDTPAPRERLIDAKTGLPVKTDITTPGSRILCSGAR 244  
 OY 193 RNSHTFCRMILV-KPLPDSSEEGHNDQEAHOKYETMOCFAVSQPSKIR----- 239  
 DB 245 R--SFCRMKCNRSVAYEDKDPSTCSKKKADKRSCTJHSTGYLKSMPPTKMGLED 301  
 OY 240 ----EEGEDLQSLICVARR---VPMK-ERPVLPSSSFTTRODLOGKITSLD-TSTMR 289  
 DB 302 NEPDNCGNL-SCIVAIRGLSHVVPQVNGEIRVKSMEYVSRAHIDKRFVVDQRAFAI 360  
 OY 290 AAMKGMGDIVRRCIOKFAHOGESVSYAKRHHNEVL--RQGLAFSQIYRFPSSLDGTLV 347  
 DB 361 LAYLP-QELIGTSCYEYHQ-----DDIGHLAECHQVQVLOTREKKT-TVCYFKIKDGSFT 414  
 OY 348 AAQTKSLIRSQTNPEPOLVISHMLHREOVNVCVNPDLTGQTMKPLNPITSSNSPAHOA 407

```

Db      415 TLRSRRESFMNPNWTKREVEIVS-----TNTVLANVLEGG--GDPTRPOLTLASP----- 460
Qy      408 LCSGNPGQDMTLSSNINPFIKNGPEKDMGMPMRGFGSGGMNHYSG---MQATTPGGSNTA 464
Db      461 -----HSDMSLPSCGCGPKRTHPTVPGIPGCTRRAGAGKIGRMIAEIMEIHRIRGS--- 512
Qy      465 LKMNPSOSSPGMNPQGPMSLSPRRHMSPG-----VASSPRTPPGQFSPAGLSHPVG 519
Db      513 -----SPSSCGSSPLNITS--TPPDASSPGCKKILNKGTPPLDPSGGLSLGQAQENP---- 562
Qy      520 CSSTGNSHSYTNSLNLALQALSEGHGVSF-----GSSLASPD 556
Db      563 -----GYEYSDSS-----SILGENPHIGIDMIDMDGGSSSPSND 596

RESULT 24
T13804
sus protein - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C:Accession: T13804
R:Treisman, J.E., Lai, Z.C., Rubin, G.M.
Development 121, 2835-2845, 1995
A:Title: Shortsigned acts in the decapentaplegic pathway in Drosophila eye development
A:Reference number: Z17767; MUID:96038094
A:Accession: T13804
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1212 <TRP>
A:Cross-References: EMBL:L42512; NID:g833709; PID:g833710; PIDN:AAC41608.1
C:Genetics:
A:Gene: shs
A:Cross-References: FlyBase:FBgn0010460

```

```

Query Match Similarity      3.5%: Score 269; DB 2: Length 1212;
Match Local Similarity      20.0%: Pred No. 1.4e-05;
Matches      251: Conservative 130; Mismatches 445; Indels 426; Gaps 53.

```

Dd	379	SVYGHFPL-----NAAGDT-----GDESADDL--ESITTD--NSR	411
Oy	818	LPLEPPTREGAPAGVDKQAI--INDLMQLTAENSPVTVPGAQKTALRISOSTENNRP	875
Dd	412	ITDLENET-PSMEDTFSKEEYYANNAL--STNPAVIPTSSQGLVV-----DPIA	461
Oy	876	GOLGRLEPNQLPLDIT-----LQSTGAGPPPPTR----NSSPYSVI----PGQ	918
Dd	462	PSLGOTI--QNVQVNSDNIIINVSGAVIPGGTKKKDDIKETHOHSERKKVKIESTEPE	519
Oy	919	MNG-----NOGMINOGNLGNSSPTMGIGNASRRPTSGEMAPQSSAAVVCATTS	970
Dd	520	KRGMRMCMDYIDHSSVNGGN-INNEKTG-----SSTSEHAIAATT--557	
Oy	971	AMNRVPOGGMIRNPASIPMRPSSOPQORLOSOYM---NIGPSELEMNN--GGPOYS	10244
Dd	558	-----DQG-----ACGV--GAGSEAPAHKHTTGSMILPTQKLNEHLNEANSTDANMNYA	604
Oy	1025	QQQARPNOTAFWPESILIPDOASFASONROFGSSPDLLCHPRAESPDSGALLDOLY	10844
Dd	605	EQQQ00000000000000FTIV-----GNALTTLKP	630
Oy	1085	LALNFGLJEIBDALDI-----PELYSQSQAVNDPEQSSODSNIMLEQKARPVPOYA	11388
Dd	631	VALRWVSRSSSVTSPNAVTEFLSPMLLAQ00000000LDVSVAN-----	675
Oy	1139	SQAOMAGSYSPMODPN-----PHTMGORPSYATLRNOPRPGRLPTGLVN---	11844
Dd	676	-----MASSPRPADBPNNMYARTAAQMOLHQTLQOKREBAMDVPFGAGGTANYONGD	730
Oy	1185	-----QPQLRLQLQHRLQAQONROPILMOISWVSUNVTJLRGPVTOAPI	12300
Dd	731	SAVGASNNNSAAAATGESQLSTSVEEQ000000P-----LSPABVLPOAAPY	778
Oy	1231	NAQMLAQORQRILLNQHRKQRMHQOOOY-----QORTLMAMGO	12684
Dd	779	FAPAAGAOSP--NEQLEQ0000000000ATSQIDGIVPDPFNQ000000TPQOSTQAQAAA	835
Oy	1269	GLNMTPSMVA-PSCGMPTMTSNPRIIPAQANAQOFEPFPNYGISGOOPP-----GFT	13164
Dd	836	AANATSAVTAPPPQOTSNTSMAAVTTGCQGYMPLLSHMTSYEDQOPNIGAAAAAAAGCT	895
Oy	1317	GATTPQSPLMSPRM-----AHMQSPMQOSQANPAYQAPSDINGWAQNMGMGSMF	13674
Dd	896	AATSVAAPOALPTQLQASASTIADDPQLLVPOQ000000HOE-----	937
Oy	1368	SQOSEPHFGQO---AN-TSMYSNNMNINVSMAINTGSHSMNMOTGQISMT	1414
Dd	938	EQQQ0PQQ0000PPLRPANIASANSNLTNI--TNTNVVATEATTNALTLTR	986

```

RESULT 25
T48814
hypothetical protein 15E6.220 [imported] - Neurospora crassa
C:Species: Neurospora crassa
C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 05-May-2000
C:Accession: T48814
R:Schulze, U.; Aign, V.; Hohnsels, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura
submitted to the Protein Sequence Database, April 2000
A:Reference number: Z24541
A:Accession: T48814
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1952 <SCH>
A:Cross-references: EMBL:AL353822; GSPDB:GN00112; NCSP:15E6.220
A:Experimental source: cosmid contlg 15E6, strain 74
C:Genetics:
A:Gene: NCSP:15E6.220
A:Map position: 2
A:Introns: 281/3

```

Query Match 3.5%; Score 269; DB 2; Length 1952;



Best Local Similarity 19.0%: Pred. No. 2.8e-05;  
Matches 292; Conservative 172; Mismatches 597; Indels 478; Gaps 65;

```

QY 67 PDKALIKETVKQIROIKEOEKAAANIDEVOKS-----DVSSP----- 105
Db 610 PDEDILLPADADHWEIYEKEEEDDDDEAKSOSQSVKPEPRPEFPPLMLDPKK 669
QY 106 -GCGVIDKDALGPMLEALDGEFFVNVLEGNVEFSENVQYLRYNOELMKNKSVYSILH 164
Db 670 RGPGRPPKNGI-----MSKREBLRKKQAMELAK 698
QY 165 VGDHTEFYKNLLPKSYNGSGWSEPP-----RRNSHTFNCMLVKPLPDSEEGHDNQ 218
Db 699 -----KNQPPQP-----PGEPPVKRKVGRRKH-----PLP-----EDAPDRP 731
QY 219 EAHQKYEIM-----QCFVNSQPKSIKEGEDQSLICVARVPMKEKRPVLPSSSFTR 273
Db 732 E-KKKYKPKRKNKEGDDSDAEKTIKEK-----RREKPKTEPLELRREDYTEE 778
QY 274 Q-----DLOGKITSL--DTSTRAAMK-----PGMEDLVRCIOKF 307
Db 779 QLOKPNKNYGLVIDEIVLSAADGLTKQIKYRIQKLYPFYFENVDTKGMESSVRINLGN 838
QY 308 HAQHEGESVYAKRHHHEVLROGLAFSQIYRFSLSDGTLVAQTKSKLIRSQTTNEPOLV 367
Db 839 DAFKNE-----ETHLMSRVPGIDIDAGKKRK-----APSPDHA 872
QY 368 ISLH-----MLHREQNY-CYMNPDLTQ-----TMGRP----- 394
Db 873 SSLNFGOHVAPQPMRPHGMYHGEHVQOSYHPEQVQOSQSYVTGQPGASQHPHOHQ 932
QY 395 -----LNPISNSPAHQALCGNFGQDWTLSNINFPINGKEQMGWGRFG-----GSG 445
Db 933 TPQGVPPQQRPRRAYQAAQTSPPAQNOVYGT---PPTARQMSGTPTAATYSSSYRSP 989
QY 446 GANHVSQMQATP---QGSNVALKMNSPQSS---PGMNGQPTSMLSPPHRMSPGVAGSPR 501
Db 990 PMPVAAQSGATPHSMARQHSLPVGSPQANGIPRVNPPATA-----NTGVPAVGAR 1043
QY 502 IPPQFSPAGSLHSPVGY-----CSTGNSHSTNSLMAQLASF-----GH 544
Db 1044 -PAQQAFTTATTAIVRLNPIVIADELISWLSFVYTEKLEYIKQTSKFPQILAMSVINR 1102
QY 545 GVSIGSSIASPD---LKMGNLQ-----NSPVMMNPPLESKM-----GSLD 581
Db 1103 GLKLTITSMIPDEESLELVYLRVEERIQGTSKSLDDLLQTLTFKATWSTIEALID 1162
QY 582 SKDCEFLYGEPESEGTGOAESCHPEQKETNDPN---LPVASSERADQOSRLHDSK 636
Db 1163 SQKAECLVLSAIDQVLGLADKITITRGTESEMRFEFNNAEKVLIPAIRMKVAEQOR 1217
QY 637 GQTLQLQLLTTKSDOMEPSLASSLSDTNKNSTGSLPSSGSGTSGTSLKEKIKILHRLQD 696
Db 1218 -----QVATTPAPVHAIR---TPGALLPANHHT----- 1242
QY 697 SSSPVDLAKTLAEATGKDLQSOESSSTAPEGSEVTIKQEPVSPKKEKALLRLILDKDQD 756
Db 1243 -----MAPNTAPSAQRIATPSAT-----QRANPPGCGNPSANPKRI---TAA 1281
QY 757 IGLPEITPKLERLDSKDPASNTKLIAMKTEKEMSEFPGQPSGEIDNLEIILDLQNS 816
Db 1282 AAPATPPALLPR---AAPISGHPRIASTVNNTNAAM---GARP----- 1318
QY 817 QLPQLPFDTRGARAGSV-----DKQATIINDLMQTLAENSPTVYPGAOKTALRISQS 868
Db 1319 -VPAAPVGPQIGVSHGAIGYATVAVSRSPVTHGVTAAPTTTPPSNLAAAVPAGSMQO 1377
QY 869 TFNNPRPG---QGRLLPNQNLPLDITLQSPGTAGPPPIRNS--SPYSVIP--OPGM-- 919
Db 1378 THSGAPAGSGNASISRAP-----STAPTAFTVTPPVSSIVYRPMASIVPTGPGLTA 1429
QY 920 --MGNOGIMGNLGNST--GMIGNSASRPTMPSGEMAPQSA-----VAVTCAA 967

```

```

Db 1430 PPTASSGASAGYARPANASTMPAPTSGQAAMTSAVPOQSVSPRPSLTTGQMGVPAIAA 1489
QY 968 TTSMNRNVQGMIRNPAS---IPMRSSQPCQRTTLOSQVWINIGPSL---EMMGGPO 1022
Db 1480 STS---RPASG--VYNPASSIASPTSKSMPSAVPTTASGAVSTSVSLAATPLPPSAPR 1544
QY 1023 Y--SQOQAPNPQATAPWPSIILPIDQASFASONROFGSSPDLLCPHPAASPSDEGALL 1080
Db 1545 YGPSNNATPTPTATAIPIRAPLIPAASSVASAPYTGQSFSPAPASVPTPTSAAS----- 1596
QY 1081 DQYLALRNFGLERIDBALGIPELVSOQAQVDPPEQSSQDSNIMLEQKAPV--PEQOYAS 1139
Db 1597 -----QGAQPLSQPVGPAPAPAISSSIGATPAASIPSSAPAL---APVTYVPQQA 1643
QY 1140 QAQMAQSVSPMDPNFTFMQORPSYALIRQPRRGLAPRTGLGYONQPNQLQLOHRLQA 1199
Db 1644 SAAARLPVTPA--PAATTIQOSVA-----QVABP---VYQSP--VQSVAAQHTSS 1688
QY 1200 QON---RQPLMNQISNVSNVNLTLRPGVP-----TQAPINAOMLAQOREIILNQHLRQ 1249
Db 1689 QASTTAHPVAQSVPRVNSNPTSAAPVAAVTAQVAAAPVTYQAPRHALSSVSQSLPQ 1748
QY 1250 ROMHQOQOVQOQRTLLMRQGLNMTPSVAPSGMAPATMSNPRIPOANAOQFPFPVYGISQ 1309
Db 1749 SVPHAAQQAHAQT-----PHSASRPVQSVPOS-----VQKRTQAVPRPSTSLTP 1795
QY 1310 QPDGFGTATTPQSPILMSPRAHQSPPMAQOSQANPAIQA----- 1349
Db 1796 TAQGPVSPAVSGSGVPAPSAQSVAPAPVASTPVPAATVAPASTVAAAPPTRTVTAAPA 1855
QY 1350 -----PSDINGMAQMGNSMFSQSPPHFGQQAQT 1381
Db 1856 ALSATINPAPYSPQPHQITQGAQQAQORPRAQAQAPAT 1894

```

## RESULT 26

30556  
aryl hydrocarbon receptor alpha chain - rainbow trout  
C:Species: Oncorhynchus mykiss (rainbow trout)  
C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 11-May-2000  
C:Accession: J30556  
R:Abnet, C.C.; Tanguey, R.L.; Hahn, M.E.; Heideman, W.; Peterson, R.E.  
J. Biol. Chem. 274, 15159-15166, 1999  
A:Title: Two forms of aryl hydrocarbon receptor type 2 in rainbow trout (Oncorhynchus  
A:Reference number: 220858; MUID:99262676  
A:Accession: J30556  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1058 <ABN>  
A:Cross-references: EMBL:AF065137; NID:g4007963; PID:g4007964; PIDN:AC95335.1

## Query Match

3.5%; Score 268.5; DB 2; Length 1058;  
Best Local Similarity 20.0%: Pred. No. 1.2e-05;  
Matches 251; Conservative 170; Mismatches 449; Indels 387; Gaps 63;

```

QY 17 KKRKECPQQLGSP-----KRNTERKRNQENKTYIELAELI-FANFDINDNFNPKPKC 70
Db 11 KKRKKPVQKTKKSPAPDVYKSNPSKRHRDLRNLGELDRITGLLPP-----EDVSRIDLK 65
QY 71 AILKETVYKQIROIKEOEKAAANIDEVOKSDVSTGQGVINDKAL-----GPMLEAL 123
Db 66 SVLRISVGYLK-VKSFETK-----MKSSVLPFGGGGLMMNGMKNATPFSEGDLLQAL 118
QY 124 DGEFFVNLEGNVVFSENVQYLRYNOELMKNKSVYSILHVGDHTEFYKNLLPKSYNG 183
Db 119 NGFVLVYTAEGHVFYSSPTIDYLGAFGSDVHVQSVFELIHTDRATF----- 166
QY 184 GSWGSEPPRRSHFTNCMLVYKPLPDSEEGHDQEAHQKETYMQCAVSPQSKKEGE 243
Db 167 -----RQDLH-----FALNPKRFDEQGDGMA--RSSDITRNIVTYNPEDLPPENS 211
QY 244 D-IQSLCLICVARVPMKEKRPVLPSESEFTTRQDLOGKITSLDTSYM----- 288

```

```

Db      212 SFLENFVC-----RFRCILDNSSGFLA-LNFQGLKFLHAQSMIGDGTSPNLTG 263
Oy      289 -----RAAMKQGMEDVLRIC-----OK 306
Db      264 FTIATPVQNPSTLEIRKTIFFQTKHKLDFTPMGVADARCVVLYGYSM-ELCHRGSSYQF 322
Oy      307 FNAHGESESVYAKRHHHEVLROGLAFSOIYRPSLSDGLVAAQTKSLIRSOITNEPOL 366
Db      323 IHA-----ADMFCADNHNVRMIKTGESGLTFTRLLQTKGCWVWVQANARLV--YKGGPRDF 376
Oy      367 VISL-----HMLHR-----EONVCYMANDLTG-----QYNGKLPNLTSSN 401
Db      377 IARQARALLNSEGEHELROKRMELPESFTTGELLYETGPTLATEPOTSPKIRKRESL 436
Oy      402 SPAHQALCSGNPGQDMTILSNINFPINGPKQKGMPCGR-FGGSGGNHY-----SGMQATP 457
Db      437 DP--QSLTGLSLMQDESITYQ-----FOERO-LPILDOAFMDSKRLTNYACMSWQSSME 466
Oy      458 PGGSNTYALKMNSPSSQSPGNNPGQPTSMLSPRHRMSPGVASPRIPPSQSPAGSLHSPV 517
Db      487 PGCPDGDGDDGDPSEVNV--QKGAIVYAMIDALEKMARL-----GDL----- 524
Oy      518 GVCSSITGNSHTYNNSSLNALQAL--SEGHCYSLGSSSLASPDLMKMLQNSPVNMMNPPLS 575
Db      525 -----CEALQGLDVEAEELMEWESAL--LRISOESNQTGGGDTSP-- 562
Oy      576 KMSGLSDKDCF-----GLYGEPEGTTGQAESCHPGEQKETNDPNLPRAVSEERADQGS 630
Db      563 ELDDITNDTDFSYVEELRFBESSEGSQNPNCST--MYNNNPN----- 603
Oy      631 RLHDSKGQTKLLQLLTTKSDOMEPPSLASSLSDTKNDSTGLSPGSGTHGT-----SLKE 665
Db      604 -----LNLFT-----EYFNNNNOD--GPPGMVSPITGVGCKPQLLD 638
Oy      686 KHKILRLDQSSPYDLAKTLAATGCKDLSQSSSTAPSEVITIKQEPVSPKKKEAL- 744
Db      639 SRSFTH-----NSPVA--SLNGCVTGN-----GPDLAGCQNAQPHQVFESTQ 660
Oy      745 -LRYLLDKDPTKDIGLEI-----PPLERLDSTKTPASNTKLIAMTEKEEMSFE 794
Db      681 RLSHFGQIPQMDLNTPTLQQLQDLNDFTPSL-LEPLSLPHSSGQNGAV-TCTNNAGS 738
Oy      795 PGQDQSELDNLEILLDLQNSQLPOLFPDTPRGAPR--GSVDKQAI-----I 840
Db      739 CAQAPRNHMSPOGITGRVHNSQPPQF-FTNHGLPATMASNGPQISVQSNHVAPLV 797
Oy      841 NDMLQTLAENSPTPYGAQKTALRISQSTFN--NPRRGQL--GRLLPNQMLPDLITLQSP 896
Db      798 DGNASMLPSPNAFVSP-----QIESSNINLSNPLPTACLOGNSAFQSLKIQRYLQWP 849
Oy      897 TGAGPPPLTINSSPYVITPQGMNGMGNOGMLGNSSTGMIGNSASRPTMP----- 950
Db      850 QNQOQLRP-----PASTI-QNGIMANGHTIPROCHSDSTQRYVPLGLIMPQPNRLXHQ 903
Oy      951 -----SGEMAQSSAARVTCATTSAMNRPVQGMIRNPAASIPMRPSSQPGQROTLQ 1003
Db      904 TQHGGLANQGPAPSSSC-----MFENISPLRPNGNSHVQDTR--LA 942
Oy      1004 SGMNNTGSPLEKMMNGPQYSSQQAAPPNQA--PWPESTILPIDQASASNNRQPFSSPD 1061
Db      943 STL-----SYCQSRMVDP--QDQSPKPGSCYQWQGPS-EPVVGTSVAVIDST--STSP 991
Oy      1062 DLICPHR-AAESPDEGALLDQLYLALRNFDGLEI-----DRAIGPELVQS 1109
Db      992 S-----RPVANTITTEGILLAMQOYLACGSGVGTQIPSLPVIDSNGILSLPPLYNGS 1044

```

RESULT 27  
 T00022  
 B120 protein - human  
 C:Species: Homo sapiens (man)  
 C:Date: 22-Jan-1999 #sequence\_revision 22-Jan-1999 #text\_change 21-Jul-2000

```

C:Accession: T00022
R:Takeuchi, T.; Chen, B.K.; Qiu, Y.; Sonobe, H.; Ohtsuki, Y.
Gene 204, 71-77, 1997
A:Title: Molecular cloning and expression of a novel human cDNA containing CAG repeat
A:Reference number: Z14059; MID:98094256
A:Accession: T00022
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1142 <TAR>
A:Cross-references: EMBL:AB001895; NID:g2588990; PIDN:BA423269.1; PID:g2588991
C:Genetics:
A:Map position: 1p35-36.1
A>Note: B120

Query Match      3.5%; Score 268; DB 2; Length 1142;
Best local similarity 19.6%; Pred. No. 1.5e-05;
Matches 262; Conservative 116; Mismatches 424; Indels 536; Gaps 61;

Oy      382 MNPDLTGQTKGKPLNPSSNS--PAHQALCSGNPGQ-----DMTILSNINFPING 429
Db      7 MRQPYRGCT-----NPSQQQGGPSPDQGHGVRGQPYGSGTQRYRMYVQGAQSAMG 61
Oy      430 -PREQMGMPKRGFGSGGGMNHVSGMQATTP--QGSNTYALKMNSP-SQSSPGMNP-GQPT 483
Db      62 LSYTQQLTPRYGQGGQPSG--YGOOGQTPRYNQDSHPHQOQPPYSSQGPSPQTPHAQPS 116
Oy      484 SMLSPRHR-----MSPGVAGSPRIPPSQSPAGSLHSPGVCSITGNSHTYNNSSLNAL 537
Db      117 YQOQPGSQPQLQSSQPPYQSPQSPRHQQSPA-----PPSQOSTTQOHP----- 162
Oy      538 QALSEGHCYSLGSSSLASPDLMKMLQNSPVNMMNPPLSKMGLSDKDCFGLYGEPEGTT 597
Db      163 -----QSQPP-----YSQP----- 171
Oy      598 GQAESCHPGEQKETNDPNLPRAVSEERADGQRLHDSKGQTKLLDITTKSQMPEPSL 657
Db      172 -QAQOS--PYQOQOPOQP-APSTLSQQAATPQSQSQSQ-----QTAYSQOQREPP- 217
Oy      658 ASSLSDTKNDSTSLSPGSGTHGTSLEKHKILRLDQ-----SSPYDLAKTLAE 709
Db      218 -----QELSDQSPGS-QASSAPMTSSKSGQEDMNLISQSRPSSLPDLGSDIDLPAGTEG 272
Oy      710 ATGKDSLQESSSTAPSEVITIKQEPVSPKKKEALLRYLLDKDPTKDIGLEITPYLERL 769
Db      273 ALSPGVSTSGISSQGEQSNAPQSPSPHNSPH-----LPGI-----RG 311
Oy      770 DSKTDPASNTKLIAMTEKEEMFBERGQ-----PSELD-----NLEILLD-----L 813
Db      312 PSPSPVGSPPASVQASRGLSPPAVPPGNOMPPRPSSQSSDSIMHPSMNOSSIAQDRGYMO 371
Oy      814 QNSQLPOLFPDTPRGAPAGSVDAKOAITNDMLQTLAENSPTPYGAQKTALRISQSTFN-- 871
Db      372 RNSQMFQ-YSSPPGSAI-----SPQQLSGGQ--IHTMGSTYQON 408
Oy      872 -----NRRPQGLRLPNQMLPDLITLQSEPTGAGPPPLTINSSPYVITPQGMNGMGMI 926
Db      409 SMGSGPGQGGQYG-----PQGYPRQPNYNALPNANYPSAGAGGINPA 452
Oy      927 GNGGNL-----GNSSTGKIGNSASRPTMP--SGEVA----- 955
Db      453 GAGGQNMHGQIPRYGTLPPRGMRSHASMGN--RPGYPNNGOYATSGWWDVSPPGGMNR 509
Oy      956 -POSSAVRVTCATTSAMNRPV-----QGM-----IRNPAASI-PMRPSSQPG 997
Db      510 KTQETAVAMHVA--NSTQNRPPGYPMNMGCGMAGTGPPYQGINSAAGMINTPQGPYRSMG 568
Oy      998 QROTLQ-----QYVNTGSEL--EKNM--GAPQYSSQQAAPPNQTAPWPESILPT 1043
Db      569 GTMANNAGMAASPENMGLDVKLTPTATKMNNAADGTPKTESKSSSTTTNEKITKL 628
Oy      1044 -----DQASFASQNRQPPGSSPDLLICHPAAESPDEGALLDQ 1083

```

Db	629	YELGGGERKMWDRYLAFLFEERKAMGTMLPANGRRPDLV	-----RL	671
OY	1084	YALARNFDGLEEDRALGIPEL	-----VSGQAVD	11133
Db	672	YVSVEKEIGLLTYVKKKKWRELATNLNVGTSSSAASLTKKXOICLAFVFECKIERGEDPP	-----	731
OY	1114	PEORSSODSNIML	-----EOKAPVFPQOYASAOQAOG	1146
Db	732	PDIFMAADSKSOPKIOPPSPAGSGMGPQTPQ	-----STSSMAEGDLKPTPTASTPHSQ	789
OY	1147		-----SYSPMDP	1154
Db	790	IPPLPGMSRNSVGIODAFNDSDSTFQKRNSMTPEPGYQPSMNTSDMGKRNSTEPNDKP	-----	849
OY	1155	-----NFTMGQRP	-----YATLRMPR	1174
Db	850	YGSMRKAPGSDPFMSQGGPNMGMDPPYSNAGRGIGLVAMGRQHNYPGCYDRVRETER	-----P	909
OY	1175	GLRPTG	-----LVQNPQ	1219
Db	910	GIGEGKMSGTGAPOSNIMPSPDGMYSRKYRPPQOQOQOQORHDSYNQEST	-----	962
OY	1220	LRRPVPQAPINQMLAQRRRELLNHLNRQKHQOQOQYQOQTILM	-----	1264
Db	963	-QSTPBGSPFPPO	-----QTTMYQOQOQONVRRPMDGTGGYPPARKHEGM	1005
OY	1265	-----MNGOGLNMTPSWAVSPGAPATMSNPRIPOANAQOFPPPYUGISQODDPGFGAT	-----	1319
Db	1006	YSVYVYTGQGLPQOQQL	-----PPAPPOYASQPOQAQPSPOQDYV	10599
OY	1320	TPQSPILSPRKAHTQSPMMQOQANPAYQAPSDINGMAQNGNGSMFSQ	-----QSPRHFGQ	1378
Db	1060	TERPPAGCPQ	-----NQPF	1111
OY	1379	ANTSMTSNMNNINVSATNT	-----	1398
Db	1112	GTMMQGRNDMTYVYANRQST	-----	1131

```

RESULT      28
S26058
Probable transforming protein (can) - human
C:Species: Homo sapiens (man)
C:Date: 13-Jan-1995 #sequence
C:Accession: S26058
R:Von Lindern, M.; Formero, M.; van Baal, S.; Jaegle, M.; de Wit, T.; Buijs, A.; Grosveld, G.
Mol. Cell. Biol. 12, 1687-1697, 1992
A:Title: The translocation (6;9), associated with a specific subtype of acute myeloid leukaemia, results in a fusion of the ABL and FIP1 genes, creating a novel ABL fusion gene
A:Reference number: S26058; MUID:92195315
A:Accession: S26058
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-2090 <YON>
A:Cross-references: EMBL:X64228; NID:q29652; PIDN:CAA45535.1; PID:q29653

```

Query Match	3.5%	Score 266.5;	DB 2;	Length 2090;
Best Local Similarity	20.7%	Pred. No. 4.1e-05;		
Matches 345; Conservative	181;	Mismatches	600;	Indels 541; Gaps 84;

[illegible]

QY	150	----	NOEJLMAKSVYSLIHVGDHTEPKNLLPSYINGSW----	SEEPARRSHFTN----	199
Dd	859	NTLANREIILNOQKRRLNHLVDLSLOOLRLTKOTSL-----	WLSLSAVPSQSSISHFSDSL	913	
QY	200	---CRMJVKPLPDEEGGHNOEA-----		220	
Dd	914	ESLGNALKTKTTIESHTKSLKRVPAKSLPMKQAOJLRNFLAKRKTPPYVTRSTAPASLSRSFL	973		
QY	221	-HOKETMOCPAVSQPKSIE-EGEDLQ-SC-----	LICVARRYPMKERP-----	VLPS	266
Dd	974	SQRYEEDLD--EVASTSVSQSLESDARTSCDDEAVYQARHRAVYKTPSTIQPSLPH	1031		
QY	267	SESEFTTROILOGTKTS--LDTSTYRANK--PGMEDLVRRCKCFHQAEGESVSTAKRHH	323		
Dd	1032	AAPFAKSHLLVHSGSPGVWGTSVATSKILIPQADSTMLATKT--VKHCAFPS-----	H	1084	
QY	324	HEVLROGLAFQOIYRFSLSGTLVYAQOTKSKLRSQTNBPOLVISLHMLHREOVCM-	382		
Dd	1085	PISAPQJLAAALRKROMASOAPAVNLTLESTL-----	KVPOV-----	NVOELK	11229
QY	383	-NPDLTGOTMKPLNPITSSNSPAHQAL--CSGNPGODMTLSNINEPLNGPEKQMDPMG	439		
Dd	1130	NNPATPSTPAWGSSV-PYSTAKTTPHVLTPYAAQAQAGSLNSLTK--PGCPPTASQ--LS	1185		
QY	440	RFGSGGNNHVSGMOATTPQGSNTALKMSPOSSFGKMPG----	OPTSMLSPIRHRMSG	495	
Dd	1186	SGDASGAKTLETAVTSTPASQOFKSPFSPSGFNFNGIITPPSPSNTFAAGATPS	1245		
QY	496	VAGSPRIPTQSPFPAAGSLHPVGVCSGTGSHSYTMSLNAJQALSEGCVSLG-	550		
Dd	1246	TKESQ--PDAFSSGG-----	GSKPSY-----	EALPESSPSGITSAANTPG	12866
QY	551	-----SLASPDLMKN--JQNSPVMMNPPLSKMSLSLSDKCFGLYGBSEGT-----	597		
Dd	1287	EPAASSSRPAVPSGTALSTTSKLETP--SKUGEL-----	LFPSLAGETLSFSGLR	1338	
QY	598	-QOAESSCHEQEKE-----	TNDPNL-----	PPAVSERADQOS	630
Dd	1339	VGOADDSRTKPKNKASSTSLTSTQPTKSGVPSGFNTAPPVLGKHTHEPVTS-----	13910		
QY	631	RLHOSKGTKLJQLTLTKSDQMEPSPLASLSTNNDSTGSLP-----	GGSGTHGSL	683	
Dd	1392	-----ATTTT--VAP-PAVISTST--AVFGSLPYVSAGSSGIVISFGTSL	14322		
QY	684	KEKHKLRLHLLQDSSSPVDLAKTLAEATGKDLQOESSSTAPASEVYIKQEPVSPKKENA	743		
Dd	1433	S-----	AGKTSBSFG--SQQJNSTVP-----	PSAP-----	14555
QY	744	LLRYLLDKDTRDIGLEIETPKL--ERLDSKTPDASNTKLIAMTKEMSEFEPDQOG	800		
Dd	1456	-----PPTTAATPLTSPPTLSFGLSSATTP--SLPMSAGSTREATSSALPEKRG	15066		
QY	801	-SELD-NLEELDLQNSQLPOLFPDTRBQAPAGSVDAKAIT-----	NJLM	844	
Dd	1507	DSEVASASASLLEEQSOAQJPOAPPT-----	SDSVKKEPVLAQPAVNSNGTAASSTLV	15611	
QY	845	QJTAENSPVTP-VEAQKTALRISQSFNNRPQOLRLLPNQNLPLDTILOSPTGAGP-	902		
Dd	1562	ALSAEAPATTVGDANTEAVPPASSFS--VEQTA-----	YTAIAAISAGTYA	16088	
QY	903	-----PIRNSPYSVIPOP-----	GMNGNGMITGNGLNGSTG--	MIGNASRP	947
Dd	1609	VETSSPTIASSTSTIAPGPSAEALAFGVTSQSSVFAQPPAASSSSAFNPQJTNMTAT--	16666		
QY	948	TMPGSENAPOSSAVRYCAFTSANNRPVQGGIRNPASTIMRPS-----	OPG	997	
Dd	1667	-----APSATPVFQGVAASTAPSLFGOOTGTASTAATPOVASSGFSFPAFTTAPG	1719		
QY	998	Q-ROTLQOVANNIIPSELE-----	MNNGGPOYSQOQ--	PPNOTAPWPEIPIPIDASPAS	1050
Dd	1720	VFGOTTTQOASVYEQGASSASAVFSFSQGFSSVPAFGQPPASTPTSTSGSVFGAASST	1779		
QY	1051	QNRQ-PRGSSPDDLCPHAPAESPSDEGALLDQJLV-ALRNEDELEIDRALGITELVSO	1108		



[illegible]

**RESULT 30**

aryl hydrocarbon receptor beta chain - rainbow trout  
C:Species: *Oncorhynchus mykiss* (rainbow trout)  
C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 11-May-2000  
C:Accession: T30557  
R:Abnet, C.C.; Tanquary, R.L.; Hahn, M.E.; Heideman, W.; Peterson, R.E.  
J. Biol. Chem. 274, 15159-15166, 1999  
A:Title: Two forms of aryl hydrocarbon receptor type 2 in rainbow trout (*Oncorhynchus mykiss*)  
A:Reference number: Z20858; MUID:99282676  
A:Accession: T30557  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1059 <AB>  
A:Cross-references: EMBL:AF065338; NID:95430724; PID:95430725; PIDN:AA095336.2

```

0Y 575 SKMSLDSKOCF-----GLXGESECTQOAESCHPGQKQNTDNP.LPRAVSEBRADQ 625
Db 564 -ELDDIMTNDIFSVYEEALREKSESSGNOPCST-----MYNNPN-----604
0Y 630 SRLHDSKGQTFKLQTLQTLTKSDOMEPSPLASLSDTNKDSGTSLPGSGTHQT-----SLK 684
Db 605 -----LNLFT-----EYFNINNQD--GFFPMVSPFTGVGCKPOLL 638
0Y 685 EKHILHRLLODSSSPYDLAKTLAATGKDLQSOESSSTAAGEVYTKOEPVSPKKENAL 744
Db 639 DSRSEIHH-----NSPYN--SLNGQYTGK-----GPDGLAGQNOAGPHQVFNST 680
0Y 745 --LRYLLDKDDTKDIDGLPEI-----TKLERLDSKTDPASTFKLMAKTEKESMF 793
Db 681 QRLSHFGPQIPQMDLNPITLQLOLQDLNDIFTPSLE-LPELIPSHSSGONGAV-TECTNMAG 738
0Y 794 EPGGQPGSELNDLEILLDQNSQLQLPDRPRGAPASVYK-----QAI 839
Db 739 SCAQAPNNHMSPPGIGIRVHSHQPPQF-FTHNGIPATMASNGPQOISTVPOSHVAPSL 797
0Y 840 INDMOLTAENSPTYPVGAQKTLARISOSTFN--NRPQOL--GRLLPNOMLPDITLOS 895
Db 798 VDGNAHSTPNAFVSP-----QIESLNLINSLPLACILOGSAPQOSIAKTORVLOW 849
0Y 896 PTGAGPPPIRNSPYSVIRPOPMGMOGNOGNIAGNSSTGMIGNSASRPTMD-----950
Db 850 PQNOOQLP-----PASTI-QINGIMANGHTFIPDCHSQSOSETRQVULFPGIMQNNRILYH 903
0Y 951 -----SGEMADPOSSAVHYTCATTTSAMNRPVOCGMRNDPASIRMPRSSQEGQOTL 1002
Db 904 QTOHGGLANGCPASSSCCMFENISPHLLNNGSHVDTBLASTLAVCSOSRMDPODOSPPK 963
0Y 1003 QSOVMNIGPSELENNMGF--QYQOQOAPRQO-----TAPPESTLIDQ 1045
Db 964 GSCYFQWBPSEPVVGTSAVJQDSTSTSPSPRLVANTITTPBCLAMQ 1011

```

## RESULT 31

DNA-binding protein Rc - mouse  
N:Alternate names: Ig kappa chain gene enhancer Recognition component  
C:Species: Mus musculus (house mouse)  
C:Date: 11-Jan-2000 #sequence\_revision 11-Jan-2000 #text\_change 18-Feb-2000  
C:Accession: T42717  
R:Wu, L.C.; Liu, Y.; Strandmann, J.; Mak, C.H.; Lee, B.; Li, Z.; Yu, C.Y.  
Genomics 35, 415-424, 1996  
A:Title: The mouse DNA binding protein Rc for the kappa B motif of transcription and  
ew family of large transcriptional proteins.  
A:Reference number: Z2236; MUID:97001141  
A:Accession: T42717  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-2282 <WU>  
A:Cross-references: EMBL:L46815; NID:g1377885; PID:g1377886; PIDN:AA840884.1  
A:Experimental source: strain BALB/c; clone T1; thymocyte, brain  
C:Genetics:  
A:Gene: Rc  
C:Function:  
A:Description: binds V(D)J recombination signal sequence and kappa B motif  
C:Superfamily: HIV-Ep2 enhancer-binding protein  
C:Keywords: DNA recombination; transcription factor

Query Match	3.4%;	Score	261.5;	DB	2;	Length	2282;
Best Local Similarity	19.68;	Pred.	No. 8.1e-05;				
Matches	331;	Conservative	187;	Mismatches	580;	Indels	589;
						Gaps	85;

```
QY      140 SENVTOYL-----RYNOELLMKSVYSIIHVGDPHE-----FYKNLLPKS   179
           ||: ||: |         |: |: |: |         |: |: |
Db       46 SESATOEIILATIOPESGPSEQETGGOOKAPRRPSIEASVHSIQLPQHPLTAFMSPGKPEH   105
QY      180 IYNGSMGSGEPPRR-----NSHTFCNRL---VKLPDSE-----                211
```

```

Db 106 LLEGSTQOLVDPMRPGSGSFEVAGSHR--OSQLPASHILPEPELPGIKVEVPRPSQV 164
QY 212 -----EGHNDQEHQKYEIMQCFNAVSPKSIKEEGDLOSLICVARRPBMK----- 259
Db 165 SLKPAEAAHKKKRPPKGYTCOYCSR-----CAKPSVLQKHINSHTG 209
QY 260 ERPLV--PSSESEFTTRDRL-----OGKITSLDTSTMRAMKPGWDLVRCIQKPHAQ 310
Db 210 ERPYPCGCGCFKTKNLKHKRSHNRKIKAGLASGSSEMYPRGLEMERIGEELPEEP 269
QY 311 HEGES-----VYAKRHHHEVL---RQGLAFQIYRFSLDGLVAQAQTKSLIRSQT- 360
Db 270 TEGESTSEETGAASPSDVLPRKHPRLSSSLY-----SSGSHGSGQERCSLSQSTG 325
QY 361 -----TNEPOLVISLHMLHRQNV---CVNMPDLTGQTMG- 392
Db 326 PSLIEDPAFPAEASHPDLKPEDTHTIKOKLALRLSERKKLLEEQTLSLPGSKGTESG 385
QY 393 -----KPLNPISNSPAHQALCSGNPQ-----DMLSSNIN-----F 425
Db 386 YFSSSEABEQOVSPPTNNAISYAEIIFGKGRIGQRTSMLASTSTQPLPLSSEDRSLV 445
QY 426 PINGPKQMGMPMGREGSGSGMNHV-----SGMQATTPOGSYALK--MNSP 470
Db 446 PLVSPRTQV-----IEHITKLITINEAVVDTSIDSVKPRRSSILTRRSVESP 493
QY 471 -----SOSSPGMNPGQPTSMLSPRHRMSPGAVGSPRIPQSPAGSLHS---PYGVC 520
Db 494 KSLIYRDSLSHGEKTKOEOBLSLSLQ-----PPSSHVPPLRLSHSMPSAAC 541
QY 521 SSTGNSHSYTNSLNLQALSEHGVSIGSSLASPDLMGNLQNSPVNMMPPRLSKMGL 580
Db 542 TISTHHTFRGS-----YSEFDHADVPS---RMTPVFTSHPRMLKRHAA 585
QY 581 DSKDCFLYGPBGTGQAESSCHPEQKETNDPNLPRAVSSRAGQSRHLSKG-QT 639
Db 586 IELPLGEY-----SSEEPG--PSSKDPSTKPSDEPPKESDILTKTKKKEFT 631
QY 640 KILQILTT-----KSDQMEPSP--LASSLDTNKSSTGSLPSSGTHGSLKEKH--- 687
Db 632 KGANYECTIGARKKKKNDNTBAHKKYCSELOITKAHSVGA-----HEYEKTQAPPEP 684
QY 688 --KILHRLDSSSPVDLAKLTAEATGKDLISOESSSTA--PGSEVITIKOEPVSPKKR-- 740
Db 685 WSGMNHKL--GATLELTPLRKRRKKSLSGDEEPPAPACPGSEYAHNRPLIGSTSPA 741
QY 741 ENALLRILDKDQTKDGLP-----EITPK---LERLDS--KTDPASNTKLI 782
Db 742 EASKASPSLEDPRASSPGLPQSELQNGRGRGECPPKFTVIQHTSSFEKSDPEQPSGL 801
QY 783 AMKTEKEMSPEDDQGSSELDNLEETLDDLOQLPOLF-----PDRRPGAPASVD 835
Db 802 EEDKPPAOFSSPPAPRGRSAHSIQPRLVQNPQVPELITVEPDRDPEPEPEPEPE 861
QY 836 KOALIT-----NDIMQLTAEKSPVTPYGAOKTALRIS-----OSTFNNRPPQOLGRL 882
Db 862 KTEEFQMPQRTQTLQALPAELP-----PKKRLRLAEMAQSSSESSFESEVP--LBR-S 913
QY 883 PNQMLPLDITLQSPGA-----GPPPIRN---SSPYSLP----- 915
Db 914 PSQSSSISLSGSSRSASAFREDHGAKAPGPFSDTRSKTLGSHMLTVPSSHPAREMRS 973
QY 916 ----QPMMGQMGIGNOG--NLGNSSTGMIGNSASRPTMSGEMAPQSSAVRYTCAMTTS 970
Db 974 ASQSPVAVPHSHMTETRSKSEYDGLSPGSLAVPAAPPAPEP---RRKCFIVRQ 1030
QY 971 A-MNRPVQGM-----IRNPASIPWRPSSQP---GQRQTOQOVMYNGPSELE 1015
Db 1031 ASLNRPEALEAVRENKRAVRMPAASKPSTKSSVPOISVGTGQ-----GPS--- 1079
QY 1016 MNMGPOYSQQAAPPNOTAPWESILPIDQASFASONRQPGSSPDLLCLCPHAESPDS 1075

```

---

```

Db 1080 ---GKSQMODRPELSSPPYTEAL-----OVFQPLGTQ-----LPPAS----- 1116
QY 1076 EGALLDQLYALRNFDGLBEIDRALGIPELVSOQAVDEPQSSQSNIMLEQAKAPFPQ 1135
Db 1117 -----LFSIQQLLPQEQ-----QSSSEFFPT 1137
QY 1136 QVASQAMAGSYS-PMQDNEFTMGQRPYSATLRMQPRGLRPTGLVONQPNQLQ-- 1192
Db 1138 Q--AMAGLSSPYSMPPLP--SLFQAP-----PLP-LQPTVL---HESQLHPQL 1180
QY 1193 LQRLDAQQNRQPLMQI-----SNVSNVNLTLRPVQVQAPINAGMLAQOREILNQHL 1247
Db 1181 LPHADIPFQQPSFLPMPCPAPSTLSGYFL-----PLQSFALQPEGI-ESH 1229
QY 1248 RQROMHQOQOVQORTLMMRGGLMTPSMVAPSGMPATMSN-----PRLPOANAQOF 1299
Db 1230 PPVK-----TSLPPLATGPG--PSSSEYSSDQLQPPVYQATSPAP 1270
QY 1300 PPPNTGISQDP-----PQFTGA--TTPQSPMSPRMAHTQSPMQSQ 1342
Db 1271 TSAPLALPACPDAWVSLVVPVRIQTHMPSYGSAMTTLISQILY-----TQSPGSPAST 1324
QY 1343 ANPAYQAPS-----DINGMAQGMGNSMFQSOOS---PPHF--GQQAATSMYSNN 1387
Db 1325 ALRKYEPSSKSMTVCEADY---YEAPGSSISKQONRGYQTPYLRVPERKGTSLSSBG 1381
QY 1388 -NNINVMATNTG-----GMSNMQMTGQ-----ISMTSVTSVS 1420
Db 1382 ILLEGSSSTASGSKRYLSPAGSLLELMTQOQKRYKKEEASKADEKLELVYCSVLLS 1441
QY 1421 TSG-----LSMCPREOVNDP---ALRGMLFPNQLPGMMIQL 1455
Db 1442 TEDRKTEKPHVGQGRREAREATLSLSD-VSDPEKLSPLSHSTLSHGTAQSALKE 1500
QY 1456 EGDTPRK 1462
Db 1501 YADPSSK 1507

```

RESULT 32  
T34513  
hypothetical protein ZK783.1 - *Caenorhabditis elegans*  
C:Species: *Caenorhabditis elegans*  
C:date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
C:accession: T34513  
R:Favell, A.; Vaudin, M.  
submitted to the EMBL Data Library, August 1994  
A:Description: The sequence of C. elegans cosmid ZK783.  
A:Reference number: Z21536  
A:Accession: T34513  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-3507 <FAV>  
A:Cross-references: EMBL:U13646; PIDN:AAC24418.1; GSFDB:GN00021; CESP:ZK783.1  
A:Experimental source: strain Bristol N2; clone ZK783  
C:Genetics:  
A:Gene: CESP:ZK783.1  
A:Map position: 3  
A:introns: 14/1; 48/2; 84/1; 196/3; 303/1; 381/1; 586/1; 605/1; 1175/3; 1207/1; 1409/3504/1

Query Match 3.4%; Score 258.5; DB 2; Length 3507;  
Best local Similarity 20.5%; Pred. No. 0.00021;  
Matches 237; Conservative 151; Mismatches 451; Indels 295; Gaps 55;

```

QY 360 TNEPOLVISLHMLHRQNVCMNPDITGQTMGKPLNPISNSPAHQALCSGNP----- 413
Db 1984 TTAEPHTTST-----SSTSTKDWMT-----SKSPENVYMSSEPEVSTSS 2025
QY 414 ----GQDMLSSNINPFGPKQMGMPMGREGSGSGMNHV---SGMQATTPOQ---GSN 462
Db 2026 SKSTTASSETTVSST-----PSSSSSEAPLRTSSPATTTETTESVSKSTTEKSSSE 2078

```



Db 1089 LSDPLSLITPOVAQEPV-----QTTISNKNVITPSPSETHPQSAVNAVQVYPMNP 1137

Qy 655 SPLASLSLDYTNKDSGLFSGSGTHGTSLKEKKILHRLQDSSPYDLAKTLAATGKD 714

Db 1138 SLVPSHSHESQVRSRTKLP-----SNFKEDD-----IRHSTCDNCKKSEKSNAD 1184

Qy 715 LSGESSSTAPGSEVYIKQVPSPKKKNMALLRYLKDQDKDGLPEITPKLERLDSKTD 774

Db 1185 ISMPSS---PASTIV-----YVQSQRMDOHLE 1209

Qy 775 PASNKLIAKKEKEKEMSF-----EPDQPGSELDNLEIILDDIQLNSQLPQ 820

Db 1210 QQRGEQLA-ATDDQDQICVVKLPHRISTNLGRPRQOQSPSGVGNVGTSHGLTNGRPP 1268

Qy 821 LPPDTRPGAPAGSVDKQALINDLMQLTA-----ENSPVTPVGAQRTALRISOSTENNPR 874

Db 1269 LPPPRK---SRAKPKKKNAFEDQMNASSGHAAYQMNSPEVSIK-----LLKQOHLN-R 1319

Qy 875 PGLGRLLPQNPLPLDITLQSPFGAGPFPPIRNSPYVIP-----QP 917

Db 1320 PNOFNNISQSQSIDL-ATLQSQVMLSP-KLQRYSEQOTMQKFSHNQKQAVHEKKNNMQ 1377

Qy 918 GMMGNQMGICNGNIGNSSTGMIGNSASRPTMSPGEMARQSSAVRYTCATTSAMNRPVQ 977

Db 1378 GSAKEPPIQHRVNLQOVPTQSSGPAHOSTI---HYAPKKQIKTL----- 1421

Qy 978 GGMINPAPASIPMRSPSGQRQTLQSQVYNTIGPSELEMMNGRQYSSQOQAPRQATPW 1037

Db 1422 -----SLQSNKTSQOHSQTTLQAPV-----EAPRQDQ---FP 1450

Qy 1038 ESILPIDAQSPASQNRQPG-----SSPDLILCPHAAESPSDCALID----- 1081

Db 1451 ENFL-EPDPSDKVTEVINNVAVASVSPDEF---RGOKQFTKEKHLIDGTSKPRPK 1505

Qy 1082 QLYLALRNPDGEEIDRALGIFELVQSQAANDPECFSSQSDSNIMLEKAVFPQQAQSA 1141

Db 1506 OSHVDRSRSGSF-VHYVGQKQOMISQOQNRNAVQVQOQO-----QPNASSL 1550

Qy 1142 QMAQGSYSPMOPRNFHTMGQR-STATLRMQRPGI-RPTGLVQONPQLRLQ----- 1194

Db 1551 STMNQYEVLD---NTNLQNPMSBQYQVMQNTQTSIVQCTSVQDGRFLQDQIQVYINQ 1607

Qy 1195 -HRLDAQNRQPLQMOISVSNVLTLPKGVPTQADINQMLAQORE----- 1241

Db 1608 THRVQYQHH--VQNHQNRN-GPTKAPRKRTKRPAPVQSRVSALHERAQIVDFAKTQPA 1664

Qy 1242 -----ILNQLRLQRQMHQOQOQVQQRITLMRGQGLMTPRSMVAPSRMPTMNPRLPQA 1294

Db 1665 DQEDQTVQQLDDEHQNYSQOQPEYQHQPLVQMOE-----QS 1700

Qy 1295 NAOQ---FEPFPNYGISQO--PDPGFTG-ATTPQSPILSPRMAHTQSPMMQSQANPAYQ 1348

Db 1701 NBOQGVQYQHNQHNQAAQDELQENQSGMGAATQOQOQCRQSLQOQLOQPRQOQMLQOQYQ 1760

Qy 1349 --APSDINGMAQGNNGSGNSMESQSPPHFGQANTSM-----YNNNNINVSM 1394

Db 1761 NVGPSHIQ---QNVNAPRQHV---QOQPMQYAAQOQOQPMQMAQQRSMQEDYVQOQGVQRTG 1815

Qy 1395 ATNTGMSG--MNQMTGQISMTSVTSVST 1421

Db 1816 ATHVILMTSPRPNHHSQANVLRAVPKQT 1844

RESULT 34

T13398

gene mastermind protein - fruit fly (*Drosophila virilis*)

C:Species: *Drosophila virilis*

C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 17-Nov-2000

C:Accession: T13398

R:Newfield, S.J. Tachida, H.; Yedvobnick, B.

J. Mol. Evol. 38, 637-641, 1994

A:Title: Drive-selection equilibrium: homopolymer evolution in the *Drosophila* gene mastermind

A:Reference number: z17850; MUID:94365848

A:Accession: J13998  
A:Status: preliminary; translated from GB/EMBL/DBD  
A:Molecule type: DNA  
A:Residues: 1-1655 <NEW>  
A:Cross-references: EMBL:M92914; MID:9157833; PID:9157834; PIDN:ACC37201.1  
C:Genetics:  
A:Cross-references: FlyBase:FBgn0013119

Query Match	3.4%;	Score 257.5;	DB 2;	Length 1655;
Best Local Similarity	20.1%;	Pred. No. 8.2e-05;		
Matches 247;	Conservative 132;	Mismatches 489;	Indels 359;	Gaps 50;

[illegible]









Db 441 IWSRDSFHSSEPPSASH-----DQLMDHLGFPDLQDPLHLHLS-PPSADHRKE 489  
QY 782 IAKTTEK--EEMSEFEP-GDPOGSE-----LDNLEETLD---DLONSOLQPLPPDR-PC 828  
Db 490 TPREECHERDELMTLDPRLPPCGSEFRGIGTIDDMPLSLDDEDELMSESAISSLSSSGRN 549  
QY 829 APAGSVDKQALINDLMQLTENSPTV-----PYGAOKTALRISQSTFNNPRGQLG 879  
Db 550 TPAVAFETRP-----TKRRKPRATKRPSEKKEASEGPEKNSAIIAALKEPAAPQOPQ 601  
QY 880 RLPPNOMLPLDITLQSTGAGPR-----PIRNSPEYS 912  
Db 602 QLOFOQNYOPSPQFOQAEYGGSLPSISASVLSASTSAAAEERSEMFTSPITVTSAPN 661  
QY 913 VI-----PQPMGMGNOGMIKNOGNIKNSGTMTIGNSASRPTMGEMAPOS 958  
Db 662 YIHQDPRHQFOQOKSSPLEDLINQDESDADDGDSRSSSTVSNSTTTTATTTSKST 721  
QY 959 SAVRYTC-----AATTSAMNR-----PYOGGMRNPASIPMRSPQOPQOTLOSQVM 1007  
Db 722 GNPLFTCEHCARQLCSMSNLKRRATCKVAASSSNSAAS--RPPSQPTPATAPAT-- 776  
QY 1008 NIGPSELMMMGPOYSQOQAPRPNQTAWPPESTLPIQASFASONR-----QPPGSS 1059  
Db 777 -----PMLQASQAPQLQAP-POS--PMETATVYTKTVPVSANTWTE 820  
QY 1060 PDDLCPHPAESPSDEGALLDOLYLALRNFDGLLEEDRALGIPELVSQOAVDPEQSS 1119  
Db 821 KAOLISKPRSSQITFSASSSMYVGDALRAQOQOKKDDQIOT-----QFOQ 867  
QY 1120 QDSNIMLEOKARVFPPOQYASQAOAGSSYSPODPNFTMGORPSYATLMBORRGLRPT 1179  
Db 868 QQQ-----QRFQHQQOQQAG-----RIPRRPPIIL 894  
QY 1180 GLVONOPNQLRLQNLHQAQONRPLMNQ-----ISVSNVNLTLRPGVTPQAP 1230  
Db 895 NOVQPRPOVQVONHONQOMLNPIRQLQSPRRPRKGLIEHKNDLVL---ITSEPL 950  
QY 1231 NAOMLAQOREILNQLRQRMHQOQVQOQRTLMMRGGL--NMTPSMAVPSGPMATMNP 1289  
Db 951 AERMDARR-----SSEGLVAVTSTPLPPIQLPQRSQAP 984  
QY 1290 RTRQANAQORFRRPNIGTISQORPDEFTGATTPQ---SPLMSPRMAATQSPMAQOOSQANPA 1346  
Db 985 -AASRQOQOOP-RVAYQVQ-----FNGRPLRPMQRLPRLPNHQQOQHMLDQSONN-- 1034  
QY 1347 YQAPSDI-----NGAOGNMGMGNSMFSSQSPRRHFGQOQANTSMYSNNMINVSMATNTGMS 1402  
Db 1035 YQOYQVQVQVQVQOQOQNLQNHQOQHQQOQOQAPRGNKSRSHSNV-----GKME 1086  
QY 1403 SMOQMTGOISMTSVTSYSTS-----GLSSMGPEQVNDPALR 1438  
Db 1087 QEAQRQSSPLDITTSVRLSIEVHNHIMKGRPLEQGOSSVDSQSTAPRSR 1137

RESULT 39  
S39161  
CREB-binding protein - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 25-Feb-1994 #sequence\_revision 10-Nov-1995 #text\_change 20-Apr-2000  
C:Accession: S39161  
R:Chr1v1a, J.C.: Kwok, R.P.S.; Lamb, N.; Hagiwara, M.; Montminy, M.R.; Goodman, R.H.  
Nature 365, 855-859, 1993  
A:Title: Phosphorylated CREB binds specifically to the nuclear protein CBP.  
A:Reference number: S39161, MID:94019866  
A:Accession: S39161  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-2441 <CHR>  
\* C:Superfamily: unassigned bromodomain proteins; bromodomain homology  
F:1112-1109/Domain: bromodomain homology <BRO>

Query Match 3.3%; Score 252; DB 2; Length 2441;  
Best local similarity 20.6%; Pred. No. 0.00026;  
Matches 248; Conservative 142; Mismatches 442; Indels 372; Gaps 61;  
QY 392 GKPLNP-----ISSNSPAHQALCSGNPGQDMTLLSSNINPFINPKEQMGMPKRGSGSGM 447  
Db 78 GSSINPGICGNVASSPVQOGL--GGOAQOQPNSTN-----MMSLGAM-GK 119  
QY 448 NHVSGMAATTPQGSNVALKMNPSOSSPGMNPQOPRSM--LSPHRMSPGVAGSPRIPRQ 506  
Db 120 SPLNGDSSTP-----NLPRQASTSGPRPPASQALNPQAOQVGLVNS----- 163  
QY 507 FSPAGSLHSPVGVCSSTGNSHSY-----TNSSLNALQALSEHGVALSGSLASPLDKMGN 561  
Db 164 -SPATISQTP-GICMANANNQTHPLGLNLSGHSILNQAQOQOQVAMNSL----- 212  
QY 562 LQNSPVNMNPPRLSKMGSLSKDCFGLYGEPSEGTGQAEBSCHPEQXETNDPNIPLAV 621  
Db 213 -----GAAGRGAGMYPRA-----PAMQAT 234  
QY 622 SSEPAD-----GQSLHDSK--GOTKILLTLTSDQMEPSPLASSLSDTKKDS 669  
Db 235 SSVLAETLQVSPOMAGHAGLNTAQAGMTKMGMTGT-----SPGQPPSQTGQOM 287  
QY 670 GSLPGSGSTHGTSLKEKHKLILHRLDQSSPYDLAKTLTAETGKDLQSESSSTAPGSEVT 729  
Db 288 GA-----TGVPNQLASKQSVNSL---PAFPDI-KNTSVTVYPMNSQLOTISGIYPIQA 338  
QY 730 IKQEPVSPKKENALLRYLIDKDDTKDGLPEITPKLERLDSKTPDASTKLIAMTEKE 789  
Db 339 IATGPADPEKRLIQOQLV-----LLLAHNRQSR 369  
QY 790 EMSEFEGDDQSELDN--LEETLDDLQNSQLPQLPPTRRPAGASVDKQALINDLMQL 846  
Db 370 EQA--NGEVRACSLPRCTKKNVLAHMTGQAP-----KACVNAACASSRQIISHWKC 421  
QY 847 TAENSPVTPYGAOKTALRISQSTFNNPRRQGLRLPNQNLPLDITLQSPGAGRPPR 906  
Db 422 TRHDCPY-----CLPKNASDKRNOQTILG----- 446  
QY 907 NSSPYVITPQGMGMQNMIGNOGNLGNSSTGMIIGNSASRPTMPSERMAPSSAVRYTCA 966  
Db 447 --SPAS-----GIQNTIGVGAQOQNAVTSL--SNRPIDP-----SSMORAUYA 486  
QY 967 AATSAMNRYOGGMIKNNPAASTIPMRSSOPGQROT--LQSOVMN--IGPSELEMMMGPOYS 1024  
Db 487 LGIRYANQ-----QTDQLOPQVGGQDAPRAPAQOKRTILNALGNMPSVAPAGITTD 538  
QY 1025 QQQAPRNPQTAWPPESTLPIQASFASONRPF--GSSPDDL--LCRHPAESPSDEGA- 1078  
Db 539 QQ--PMLTIS--ESALP--TSLGATN-PLMNDGSSNSGNTISLTIPAPRPSSTGYR 588  
QY 1079 -----LDDQLYLAL-RNFDGLEEDRALGIPELVSQOAVDPEQSSODSN 1123  
Db 589 KGNHEHVTQDLRSHLVAKLVQALFPPRDPALKDRRR--ENLVAIVAKKVEGDYIESANSR 646  
QY 1124 ----IMLEOKARVFPPOQYASQAOAGSSYSPODPNFTMGORPSYATLMBORRGLRPT 1179  
Db 647 DEYVHLLAEKI-----YKQKLEEKRRTRRLIKQGI--LGNDPALPASAQAP-PYIPRA 697  
QY 1180 GLVONOPNQLRLQNLHQAQONRPLMNQISNVSNVNLTLR----- 1221  
Db 698 QSVRRPMPGRLPLRV-NRMQVSQ-----MNSFNPNLSIGNLQDLPAPMGPRASPMNHSVQM 752  
QY 1222 -----PGV---PTQAPINAQMLAQOREILNQLHQAQOQVQOQRTLMNR-----GO 1268  
Db 753 NSMASVPGMAISPSRMPQPRNMGMTHANNIMAPQONQFLPNQNPSSSGAASVNSVGM 812  
QY 1269 GLMWTSPMVPASGMP--ATMSPN---RTPQANAQORFRRPYVSGSQOP----- 1311  
Db 813 GQPAQAGVSGQGEQGAALPNPLMLAPQAS--QLPCPP-----VTSQSLPLHTPPASTAAG 867



**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 7, 2002, 10:38:29 ; Search time 16.67 Seconds  
(without alignments)  
3400.445 Million cell updates/sec

Title: US-09-842-256-2  
Perfect score: 7631  
Sequence: 1 MSGMGENTSDPSRAETRRK.....NQLPGMDIKQEGDTTRK 1464

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listed first 45 summaries

Database: SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7626	99.9	1464	1	NC02_HUMAN
2	7216	94.6	1462	2	NC02_MOUSE
3	7160.5	93.8	1465	1	NC02_RAT
4	5276.5	69.1	1516	1	NC02_XENLA
5	328	4.3	1507	1	STMA_DROME
6	317	4.2	846	1	CLOC_HUMAN
7	310.5	4.1	644	1	ARNT_DROME
8	309.5	4.1	855	1	CLOC_MOUSE
9	289	3.8	816	1	NP2A_MOUSE
10	289	3.8	1531	1	NP2A_MOUSE
11	284.5	3.7	1902	1	NP2A_MOUSE
12	284	3.7	874	1	NP2A_MOUSE
13	278.5	3.6	870	1	NP2A_MOUSE
14	277	3.6	2414	1	NP2A_MOUSE
15	273.5	3.6	766	1	NP2A_MOUSE
16	270	3.5	1589	1	NP2A_MOUSE
17	268.5	3.5	826	1	NP2A_MOUSE
18	268.5	3.5	3726	1	NP2A_MOUSE
19	267	3.5	583	1	NP2A_MOUSE
20	266.5	3.5	2090	1	NP2A_MOUSE
21	266	3.5	765	1	NP2A_MOUSE
22	266	3.5	1211	1	NP2A_MOUSE
23	262	3.4	706	1	NP2A_MOUSE
24	262	3.4	822	1	NP2A_MOUSE
25	260.5	3.4	712	1	NP2A_MOUSE
26	259	3.4	824	1	NP2A_MOUSE
27	258.5	3.4	3969	1	NP2A_MOUSE
28	258	3.4	2004	1	NP2A_MOUSE
29	252	3.3	2441	1	NP2A_MOUSE
30	251	3.3	1794	1	NP2A_MOUSE
31	250	3.3	1596	1	NP2A_MOUSE
32	249.5	3.3	3703	1	NP2A_MOUSE
33	249	3.3	1023	1	NP2A_MOUSE

34	248	3.2	790	1	ARNT_RABIT	002748 Oryctolagus
35	248	3.2	2175	1	HMCU_DROME	P10160 drosophila
36	248	3.2	3828	1	TRX_DROME	024742 drosophila
37	243.5	3.2	958	1	TRH_DROME	024119 drosophila
38	243	3.2	2442	1	CBP_HUMAN	092793 homo sapien
39	243	3.2	3866	1	HRX_MOUSE	P35200 mus musculu
40	241.5	3.2	3726	1	TRX_DROME	P30659 drosophila
41	241	3.2	1603	1	PSC_DROME	P35820 drosophila
42	239	3.1	673	1	STM_DROME	P05709 drosophila
43	237	3.1	413	1	CYCL_DROME	061734 drosophila
44	236	3.1	1403	1	PRO_DROME	P29617 drosophila
45	234.5	3.1	1210	1	AF4_HUMAN	P51825 homo sapien

## ALIGNMENTS

RESULT	ID	NC02_HUMAN	STANDARD	PRT: 1464 AA.
AC	015596	NC02_HUMAN	STANDARD	PRT: 1464 AA.
DT	16-OCT-2001	(Rel. 40, Last sequence update)		
DT	16-OCT-2001	(Rel. 40, Last annotation update)		
DE	Nuclear receptor coactivator 2 (Transcriptional intermediary factor 2).			
DE	NC02 OR TIF2.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	Tissue-Placenta;			
RX	MEDLINE=96312964; PubMed=8670870;			
RA	Voegel J.J., Heine M.J.S., Zechel C., Chambon P., Gronemeyer H.;			
RT	"TIF2, a 160 kDa transcriptional mediator for the ligand-dependent activation function AF-2 of nuclear receptors.";			
RL	EMBO J. 15:3667-3675(1996).			
RN	[2]			
RP	CHARACTERIZATION.			
RX	MEDLINE=98094372; PubMed=9430642;			
RA	Voegel J.J., Heine M.J.S., Tini M., Vivat V., Chambon P., Gronemeyer H.;			
RT	"The coactivator TIF2 contains three nuclear receptor-binding motifs and mediates transactivation through CBP, binding-dependent and -independent pathways.";			
RL	EMBO J. 17:507-519(1998).			
CC	- FUNCTION: TRANSCRIPTIONAL COACTIVATOR FOR STEROID RECEPTORS AND NUCLEAR RECEPTORS. COACTIVATOR OF THE STEROID BINDING DOMAIN (AF-2) BUT NOT OF THE MODULATING N-TERMINAL DOMAIN (AF-1).			
CC	- SUBCELLULAR LOCATION: Nuclear.			
CC	- SIMILARITY: CONTAINS 1 PAS (PER-ARNT-SIM) DIMERIZATION DOMAIN.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).			
CC	EMBL: X97674; CAA66263.1;			
DR	TRANSFAC: T02483;			
DR	MIM: 601993;			
DR	InterPro: IPR001974; Gelsolin.			
DR	InterPro: IPR001092; HLH_dim.			
DR	InterPro: IPR001610; PAC.			
DR	InterPro: IPR000014; PAS.			
DR	Pfam: PF00626; Gelsolin; 1.			
DR	Pfam: PF00010; HLH; 1.			
DR	Pfam: PF00089; PAS; 1.			
DR	SMART: SM00353; HLH; 1.			

DR SMART: SM00086; PAC: 1.  
 DR SMART: SM00091; PAS: 1.  
 DR PROSITE: PSS0112; PAS: 1.  
 KW Transcription regulation; Activator; Nuclear protein.  
 FT DOMAIN 119 183 PAS.  
 FT DOMAIN 1254 1260 POLY-GLN.  
 SQ SEQUENCE 1464 AA; 159156 MW; 0A61AA5D1878304B CRC64;

Query Match 99.9%; Score 7626; DB 1; Length 1464;  
 Best Local Similarity 99.9%; Pred. No. 0;  
 Matches 1463; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

QY 1 MSGGENTSPSPRAETKREKCEPQGLSPKRNTERKRNREKNTYIELAEIIFANNDI 60
DB 1 MSGGENTSPSPRAETKREKCEPQGLSPKRNTERKRNREKNTYIELAEIIFANNDI 60
QY 61 DNFNFKPKCAILKETKQKROIKEDEKAAANIDEVOKSDVSTGGVTDKDALGPMML 120
DB 61 DNFNFKPKCAILKETKQKROIKEDEKAAANIDEVOKSDVSTGGVTDKDALGPMML 120
QY 121 EALDGEFFVNLGNNVFSSEVNTQYLYRNOBELMKNKSVYSLHVGDTHEFVKMLPKSI 180
DB 121 EALDGEFFVNLGNNVFSSEVNTQYLYRNOBELMKNKSVYSLHVGDTHEFVKMLPKSI 180
QY 181 VNGSGSGEPPRRNSHTFNCRLMVLKPLDSEEGHNOEAHQKYEIMQCAVSPKSIKE 240
DB 181 VNGSGSGEPPRRNSHTFNCRLMVLKPLDSEEGHNOEAHQKYEIMQCAVSPKSIKE 240
QY 241 EGEDLOSLICIVARVPMKREPLPSESEFTTODLOGKITSLDTSIRAMKPGMEDLV 300
DB 241 EGEDLOSLICIVARVPMKREPLPSESEFTTODLOGKITSLDTSIRAMKPGMEDLV 300
QY 301 RRCIOKFHAQHEBESVYAKRHHEVLROGLAFSQYIRFSLSDGTLLVAQTKSLRSQT 360
DB 301 RRCIOKFHAQHEBESVYAKRHHEVLROGLAFSQYIRFSLSDGTLLVAQTKSLRSQT 360
QY 361 TNEPOLYISLHMLHREONVCMNPDLTGOMGKPLNPISNSPAHQALCGNNGQDMTSL 420
DB 361 TNEPOLYISLHMLHREONVCMNPDLTGOMGKPLNPISNSPAHQALCGNNGQDMTSL 420
QY 421 SNINFPINGKEQMGPMGRFGSGGNHVSQGNATTPQGSNALKMNSPQSGNPNP 480
DB 421 SNINFPINGKEQMGPMGRFGSGGNHVSQGNATTPQGSNALKMNSPQSGNPNP 480
QY 481 OPTSMLSPPRRMSPVAGSRIPRPSQSPAGSLHSPVGVCSSTGNSHSTYNSLALQAL 540
DB 481 OPTSMLSPPRRMSPVAGSRIPRPSQSPAGSLHSPVGVCSSTGNSHSTYNSLALQAL 540
QY 541 SEGHGVSLSLASPLDKMGNLQNSPVNNPPLSLKMGSLDSKDCGIVGEPESEGTGQA 600
DB 541 SEGHGVSLSLASPLDKMGNLQNSPVNNPPLSLKMGSLDSKDCGIVGEPESEGTGQA 600
QY 601 ESSCHPGEQKETNDPMLPPAVSSERADGQSLHDSKGQTKLQDLTTKSDQMPSPGLAS 660
DB 601 ESSCHPGEQKETNDPMLPPAVSSERADGQSLHDSKGQTKLQDLTTKSDQMPSPGLAS 660
QY 661 LSDTNDSTGSLRPGSGSTGTSLKEKHKILHRLQDSSSPVDLAKITAEATGKDLQESS 720
DB 661 LSDTNDSTGSLRPGSGSTGTSLKEKHKILHRLQDSSSPVDLAKITAEATGKDLQESS 720
QY 721 STAPGSEVITKOEVPSPKKENALLRYLDKDDTKDIGPETPKLERLDSKTDPSANRK 780
DB 721 STAPGSEVITKOEVPSPKKENALLRYLDKDDTKDIGPETPKLERLDSKTDPSANRK 780
QY 781 LIAMTKEKMSFEPDQSGSELDNLEILLDLQNSQLPQLPDPTRPGAPAGSVDAQIIT 840
DB 781 LIAMTKEKMSFEPDQSGSELDNLEILLDLQNSQLPQLPDPTRPGAPAGSVDAQIIT 840
QY 841 NDIMOLTEKNSVTPVGAQKTAIRISQSFNNPRPGQLRLPNONLPLDITLQSPGTAG 900
DB 841 NDIMOLTEKNSVTPVGAQKTAIRISQSFNNPRPGQLRLPNONLPLDITLQSPGTAG 900

```

```

QY 901 PEPPIRNSPYSVIPQPMGNGQMGIGNQNLGNSSTGMIGNSASRPTMPSEGAPOSSA 960
DB 901 PEPPIRNSPYSVIPQPMGNGQMGIGNQNLGNSSTGMIGNSASRPTMPSEGAPOSSA 960
QY 961 VRVTCATTSAMNRPVGGMRNPAASIPMRPSSQPOQRQTLQSVANNIGSELEMMNG 1020
DB 961 VRVTCATTSAMNRPVGGMRNPAASIPMRPSSQPOQRQTLQSVANNIGSELEMMNG 1020
QY 1021 POYSQOQAPRNOQAPWPESTILPIDQASFASQNRQPFSSPDLLCPHAAESPDEGALL 1080
DB 1021 POYSQOQAPRNOQAPWPESTILPIDQASFASQNRQPFSSPDLLCPHAAESPDEGALL 1080
QY 1081 DQLYLALRNFDGEIDRALGIPELVSOQAVDEQFSSODSNIMLEQKAPVFPQYASQ 1140
DB 1081 DQLYLALRNFDGEIDRALGIPELVSOQAVDEQFSSODSNIMLEQKAPVFPQYASQ 1140
QY 1141 AQMAQGSYSPMODNFTHTMQRPSTAYLRMQPRGLRPTGLVONQPNQLRLQHLRLQAO 1200
DB 1141 AQMAQGSYSPMODNFTHTMQRPSTAYLRMQPRGLRPTGLVONQPNQLRLQHLRLQAO 1200
QY 1201 QNRQPLMNOISNVSNVLTLPQVPTQAPINAQMLAQRORREILNQHRLQOMHQOQVOQ 1260
DB 1201 QNRQPLMNOISNVSNVLTLPQVPTQAPINAQMLAQRORREILNQHRLQOMHQOQVOQ 1260
QY 1261 RTLMKRGQGLNMTPSVYAPSGMPATMSNPRIPOANAQOFPFPNYGISQPDGFTGATT 1320
DB 1261 RTLMKRGQGLNMTPSVYAPSGMPATMSNPRIPOANAQOFPFPNYGISQPDGFTGATT 1320
QY 1321 POSPLMSPRRAHPOSPMOQSOANPAYQAPSDINGNAQGMGNSMFSQSPHFGQAN 1380
DB 1321 POSPLMSPRRAHPOSPMOQSOANPAYQAPSDINGNAQGMGNSMFSQSPHFGQAN 1380
QY 1381 TSMYSNNMNIIVSMATNTGQSSMNOQTQISMTSVTSYSTGLSSMGPCQVNDPALRG 1440
DB 1381 TSMYSNNMNIIVSMATNTGQSSMNOQTQISMTSVTSYSTGLSSMGPCQVNDPALRG 1440
QY 1441 NLEPNQLPGMDMTKQEGDITTRKYC 1464
DB 1441 NLEPNQLPGMDMTKQEGDITTRKYC 1464

```

RESULT 2  
 NCO2\_MOUSE  
 ID NCO2\_MOUSE STANDARD; PRT: 1462 AA.  
 AC Q61026; P97759; 009001;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Nuclear receptor coactivator 2 (transcriptional intermediary factor 2)  
 DE (Glucocorticoid receptor-interacting protein 1) (GRIP-1).  
 GN NCOA2 OR TIF2 OR GRIP1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sclirognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ICR; TISSUE=Brain;  
 RX MEDLINE=97265407; PubMed=9111344;  
 RA Hong H., Kohli K., Garabedian M.J., Stallcup M.R.;  
 RT "GRIP1, a transcriptional coactivator for the AF-2 transactivation  
 domain of steroid, thyroid, retinoid, and vitamin D receptors.";  
 RL Mol. Cell. Biol. 17:2735-2744(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97336097; PubMed=9192892;  
 RA Torchia J., Rose D.W., Inostroza J., Kamei Y., Westin S., Glass C.K.,  
 RA Rosenfeld M.G.;  
 RT "The transcriptional co-activator p/CIP binds CBP and mediates  
 RT nuclear-receptor function.";  
 RL Nature 387:677-684(1997).  
 RN [3]  
 RP SEQUENCE OF 322-1119 FROM N.A.



CC STRAIN-ICR; TISSUE=Embryo;  
 RA MEDLINE=96209838; PubMed=8643509;  
 RX Hong H., Kohli K., Trivedi A., Johnson D.L., Stallcup M.R.;  
 RT "GRIPL, a novel mouse protein that serves as a transcriptional  
 RT coactivator in yeast for the hormone binding domains of steroid  
 RT receptors.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 93:4948-4952(1996) **★**  
 CC -1- FUNCTION: TRANSCRIPTIONAL COACTIVATOR FOR STEROID RECEPTORS AND  
 CC NUCLEAR RECEPTORS. COACTIVATOR OF THE STEROID BINDING DOMAIN  
 CC (AF-2) BUT NOT OF THE MODULATING N-TERMINAL DOMAIN (AF-1).  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- TISSUE SPECIFICITY: UBIQUITOUS.  
 CC -1- SIMILARITY: CONTAINS 1 PAS (PER-ARNT-SIM) DIMERIZATION DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: U39060; AAC53151.1; -;  
 DR EMBL: AF000582; AAB61575.1; -;  
 DR TRANSFAC: T02482; -;  
 DR MGD: MGI:1276533; Nco2;  
 DR InterPro: IPR001974; Gelsolin.  
 DR InterPro: IPR001092; HLH\_dim.  
 DR InterPro: IPR001610; PAC.  
 DR InterPro: IPR000014; PAS.  
 DR Pfam: PF00626; Gelsolin; 1.  
 DR Pfam: PF00010; HLH; 1.  
 DR Pfam: PF00989; PAS; 1.  
 DR SMART: SM00353; HLH; 1.  
 DR SMART: SM00086; PAC; 1.  
 DR SMART: SM00091; PAS; 1.  
 DR PROSITE: PS50112; PAS; 1.  
 DR KW Transcription regulation; Activator; Nuclear protein.  
 FT DOMAIN 119 183  
 FT CONFLICT 51 51 D -> E (IN REF. 2).  
 FT CONFLICT 140 141 SE -> FR (IN REF. 2).  
 FT CONFLICT 194 194 T -> S (IN REF. 2).  
 FT CONFLICT 251 320 CVARVPKMERPTLPSSSEFTTRDLOGKITSLDTSTWRAA  
 FT MKPMEDLVRCIOKFTFHOGESLSTYAK -> VMHEDPHE  
 FT GKTNSSLIRKLYHPGPRRODHFTHGHESRHEHSGLRGSG  
 FT KDAFRSTSHMKSGSLYMPR (IN REF. 2).  
 FT G -> S (IN REF. 2).  
 FT S -> N (IN REF. 2).  
 FT E -> K (IN REF. 2).  
 FT EE -> KK (IN REF. 2).  
 FT R -> C (IN REF. 2).  
 FT T -> S (IN REF. 2).  
 FT N -> Y (IN REF. 2).  
 FT PAVRTCAATGAMNRPVQGMIR -> TSCSESTLVLLPLV  
 FT PRDOSKEARG (IN REF. 2).  
 FT R -> G (IN REF. 2).  
 FT P -> L (IN REF. 2).  
 FT C -> G (IN REF. 2).  
 FT P -> L (IN REF. 2).  
 FT SEQUENCE 1462 AA: 158511 MW: ACAD18979FCDCAB5 CRC64;  
 Query Match 94.68; Score 7216; DB 1; Length 1462;  
 Best Local Similarity 94.18; Pred. No. 0;  
 Matches 1377; Conservative 43; Mismatches 42; Indels 2; Gaps 1;

QY 121 EALDGFEEVNLGNVVESENVTOLRYNOBELMKNKSYSTLHYGDHTEFYKNLLPSKI 180  
 DB 121 EALDGFEEVNLGNVVESENVTOLRYNOBELMKNKSYSTLHYGDHTEFYKNLLPSM 180  
 QY 181 VNGSGWSEPPRRNSHTFNCRLIVPLPDSEEGHNDONAHOKYETMOCFAVSOPKSIKE 240  
 DB 181 VNGSGWSEPPRRNSHTFNCRLIVPLPDSEEGHNDONAHOKYETMOCFAVSOPKSIKE 240  
 QY 241 BEEDLQSLICVARRVPKMERPLDSSSEFTTRDLOGKITSLDTSTWRAAMPQMEDLY 300  
 DB 241 BEEDLQSLICVARRVPKMERPLDSSSEFTTRDLOGKITSLDTSTWRAAMPQMEDLY 300  
 QY 301 RRCIOKFAHOGESVYAKRHHHEVLRGLAFSOLYRPSLSDGLTYAOTSKILRSOT 360  
 DB 301 RRCIOKFAHOGESVYAKRHHHEVLRGLAFSOLYRPSLSDGLTYAOTSKILRSOT 360  
 QY 361 TNEPOLVLSLHMLHREOVNVCNPNPDLOGTMKPLNPISNSPAHQALCSGNPGODMTLS 420  
 DB 361 TNEPOLVLSLHMLHREOVNVCNPNPDLOGTMKPLNPISNSPAHQALCSGNPGODMTLG 420  
 QY 421 SNINFPINGPKQMGMPKRGFGSGGMNHVSGMQATTPOGSNYALKMNSPQSSPGMNP 480  
 DB 421 SNINFPINGPKQMGMPKRGFGSGGMNHVSGMQATTPOGSNYALKMNSPQSSPGMNP 480  
 QY 481 OPTSLSPRHMSPGVAGSPRIIPSOFPAGSLHSPVGCSTGNSHTNSLNAAL 540  
 DB 481 QASVSLSPRHMSPGVAGSPRIIPSOFPAGSLHSPVGCSTGNSHTNSLNAAL 540  
 QY 541 SEGHGVSLGSSLASPLDKMGNLQNSPVNMPPLPKMGSIDKDCDFGLYGESEEGTGOA 600  
 DB 541 SEGHGVSLGSSLASPLDKMGNLQNSPVNMPPLPKMGSIDKDCDFGLYGESEEGTGOA 600  
 QY 601 ESSCHPGEKRETNPNLPPAVSSEKADQSRUHSKSGOTKLLQLITTSXDOMEPSPASS 660  
 DB 601 EASCHPGEKRETNPNLPPAVSSEKADQSRUHSKSGOTKLLQLITTSXDOMEPSPASS 660  
 QY 661 LSDTKKDSGSLPGSGSTGTSLKKEKHLHLLDSSSPVDLAKTLTVAETKELSOSS 720  
 DB 661 LSDTKKDSGSLPGSGSTGTSLKKEKHLHLLDSSSPVDLAKTLTVAETKELSOSS 720  
 QY 721 STAPGSEYTIKQEPVSPKKENALLRYLIDKDDYDIDGIPETPKLERLSDTDPASNTK 780  
 DB 721 STAPGSEYTIKQEPVSPKKENALLRYLIDKDDYDIDGIPETPKLERLSDTDPASNTK 780  
 QY 781 LIAMTEKEKMSFEFGDQSGSELNLELIDLOLQNSQLPQLFPDTRPGAPAGSVUKOAI 840  
 DB 781 LIAMTEKEKMSFEFGDQSGSELNLELIDLOLQNSQLPQLFPDTRPGAPAGSVUKOAI 840  
 QY 841 NDLMQUTAEKSPVTVGAOKTALRISOSTFNNPRRGOLGRLLPNONLPDITLOSPTGAG 900  
 DB 841 NDLMQUTAEKSPVTVGAOKTALRISOSTFNNPRRGOLGRLLPNONLPDITLOSPTGAG 900  
 QY 901 PEPPIRNSSPSVIIPQPMGMNGMIGNOGNGNSSTGMIGNSASRPTMPSEMAPOSSA 960  
 DB 901 PEPPIRNSSPSVIIPQPMGMNGMIGNOGNGNSSTGMIGNSASRPTMPSEMAPOSSA 960  
 QY 961 VRYTCAATTSAAMNRPVQGMIRNPAASIPMRPSSQPGQRTLOSOVNINIGSELEMNNG 1020  
 DB 961 VRYTCAATTSAAMNRPVQGMIRNPAASIPMRPSSQPGQRTLOSOVNINIGSELEMNNG 1020  
 QY 1021 POYSOQOAPPNQTAWPESILPIDQASASQORPFGSSPDLLCPHPAASPSDEGALL 1080  
 DB 1021 POYSOQOAPPNQTAWPESILPIDQASASQORPFGSSPDLLCPHPAASPSDEGALL 1080  
 QY 1081 DQLYLALNFDGLIEIDBALGIPELVSOQAADVPOFSQSDSNIMLEOKAPVFPQOYASO 1140  
 DB 1081 DQLYLALNFDGLIEIDBALGIPELVSOQAADVPOFSQSDSNIMLEOKAPVFPQOYASO 1140  
 QY 1141 AQMAQGSYSPMODPNFHTMGORPSTATLRMOPRPLRPTGLVQONPNOLRLQLOAHRLOAO 1200  
 DB 1141 AQMAQGSYSPMODPNFHTMGORPSTATLRMOPRPLRPTGLVQONPNOLRLQLOAHRLOAO 1200

QY 1201 QNRPLMNQISVSNVNLTRPGVPTQAPINAGMLAQRORELLINQHLQRQMHQOQOQVQ 1260  
 DB 1201 QNRQPLMNQISSVSNVNLTRPGVPTQAPINAGMLAQRORELLINQHLQRQMHQOQOQVQ 1258  
 QY 1261 RTLMHMGQGLNMTPSWVAPSGMATNSRIRIQANQOQFPFPYNGISQOPDQFGTAT 1320  
 DB 1259 RTLMHMGQGLNMTPSWVAPSGMATNSRIRIQANQOQFPFPYNGISQOPDQFGTAT 1318  
 QY 1321 POSPLMSPRRAHQPSPMAQSOANPAYQAPSDINGNAGQNGNSMFSQOQSPHFGQAN 1380  
 DB 1319 POSPLMSPRRAHQPSPMAQSOANPAYQAPSDINGNAGQNGNSMFSQOQSPHFGQAN 1378  
 QY 1381 TSWYNNMNINVSMAITGTGSSMNQMTGQISMTSVTSSTGLSSMGPEQVNDPALRG 1440  
 DB 1379 TSWYNNMNINVSMAITGTGSSMNQMTGQISMTSVTSSTGLSSMGPEQVNDPALRG 1438  
 QY 1441 NLEPNOLPGMDMKOEEDTTRKXC 1464  
 DB 1439 NLEPNOLPGMDMKOEEDTTRKXC 1462

RESULT 3  
 NCO2\_RAT STANDARD: PRT: 1465 AA.  
 AC Q9W019;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DE 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Nuclear receptor coactivator 2 (Transcriptional intermediary factor 2).  
 GN NCOA2 OR TIF2.  
 OS Rattus norvegicus (Rat).  
 OC Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxId=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RX MEDLINE=98414616; PubMed=9742117;  
 RA Leers J., Treuter E., Gustafson J. -A.;  
 RT Mechanistic principles in Nr box-dependent interaction between nuclear hormone receptors and the coactivator TIF2.";  
 RL Mol. Cell. Biol. 18:6001-6013(1998).  
 CC -1- FUNCTION: TRANSCRIPTIONAL COACTIVATOR FOR STEROID RECEPTORS AND NUCLEAR RECEPTORS. COACTIVATOR OF THE STEROID BINDING DOMAIN (AF-2) BUT NOT OF THE MODULATING N-TERMINAL DOMAIN (AF-1).  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- SIMILARITY: CONTAINS 1 PAS (PER-ARNT-SIM) DIMERIZATION DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).  
 CC  
 DR EMBL: AF136943; AAD24587.1; -  
 DR InterPro: IPR001092; HLH\_dim.  
 DR InterPro: IPR001610; PAC.  
 DR InterPro: IPR000014; PAS.  
 DR Pfam: PF00989; PAS.1.  
 DR SMART: SM00353; HLH.1.  
 DR SMART: SM00086; PAC.1.  
 DR SMART: SM00091; PAS.1.  
 DR PROSITE: PSS0112; PAS.1.  
 KM Transcription regulation; Activator; Nuclear protein.  
 FT DOMAIN 119 183 PAS.  
 FT DOMAIN 1254 1260 POLY-QLN.  
 SQ SEQUENCE 1465 AA; 159435 MW; 36625B573EB0B39C CRC64;

Query Match 93.8%; Score 7160.5; DB 1; Length 1465;

Best Local Similarity 92.8%; Pred. No. 0;  
 Matches 1359; Conservative 54; Mismatches 51; Indels 1; Gaps 1;

QY 1 MSGGENTSDPSRAEPKREKCEPDQIGSPKRNTERKNEDEKYTELELLIFANNDI 60  
 DB 1 MSGGENTSDPSRAETKREKCEPDQIGSPKRSKTERKNEDEKYTELELLIFANNDI 60  
 QY 61 DNFNFKPKCAILKETVYKQIRQIKEDKAAAANIDEVQKSDVSTGGVNDKALGPM 120  
 DB 61 DNFNFKPKCAILKETVYKQIRQIKEDKAAAANIDEVQKSDVSTGGVNDKALGPM 120  
 QY 121 EALDGEFFVNLGNVVFSENVYQYLRVQBELMANKSVYSLHVGDTHEFVKNLPKST 180  
 DB 121 EALDGEFFVNLGNVVFSENVYQYLRVQBELMANKSVYSLHVGDTHEFVKNLPKSM 180  
 QY 181 VNGSWSGEPBRNSHTFNCRLVPLPDSEEGHNDQEAHQYETMQCFVAVSQPSIKE 240  
 DB 181 VNGSWSGEPBRNSHTFNCRLVPLPDSEEGHNDQEAHQYETMQCFVAVSQPSIKE 240  
 QY 241 EGEDLOSLICVARYPMKEPVLPSSESFTTRQDLOGKITSLDTSTRAAMPGWEDLV 300  
 DB 241 EGEDKQCLICVARYPMKEPVLPSSESFTTRQDLOGKITSLDTSTRAAMPGWEDLV 300  
 QY 301 RCIQKFAHGEBSVYAKRHHEVLRGLAFSQIYRESLSGTLVAAOTKSLIRSQ 360  
 DB 301 RCIQKHTQHEGESLYAKRHHEVLRGLAFSQIYRESLSGTLVAAOTKSLIRSQ 360  
 QY 361 TNEPOLYISLHMLHREBNVCVMPNDLGTGTMGPPLNFISSNSPAHQAICSGNGODP 420  
 DB 361 TNEPOLYISLHMLHREBNVCVMPNDLGTGTMGPPLNFISSNSPAHQAICSGNGODP 420  
 QY 421 SNINFPINGPKMGMPGRFGSGGNHVSQMATTPQGSNYALKNMSPSQSPGMPNG 480  
 DB 421 SNINFPINGPKMGMPGRFGSGGNHVSQMATTPQGSNYALKNMSPSQSPGMPNG 480  
 QY 481 QPTSMLSPRHRMSPVAGSRIRIPQSPAGSLHSPVGVCSSTGNSTYNSLSMALQAL 540  
 DB 481 QPSSVYLSPRHRMSPVAGSRIRIPQSPAGSLHSPVGVCSSTGNSTYNSLSMALQAL 540  
 QY 541 SEGHGVSLSGLASPLDKMKNLQNSPVNMNMPPLSKKGSLSKSDKDCFLYESEGTGGA 600  
 DB 541 SEGHGVSLSGLASPLDKMKNLQNSPVNMNMPPLSKKGSLSKSDKDCFLYESEGTGGA 600  
 QY 601 ESSCHPEQKRETPNLPVAVSSERADQSLRDSKQTKLQDLTTKSDQMPSPASS 660  
 DB 601 QASCHPEQKRPNDSPVAVSSERADQSLRDSKQTKLQDLTTKSDQMPSPASS 660  
 QY 661 LSDTNKDSLSLGSSTHGTSLKEKHKILHRLLODSSPYDLAKLAEATGDLQESS 720  
 DB 661 LSDTNKDSLSLGSSTHGTSLKEKHKILHRLLODSSPYDLAKLAEATGDLQESS 720  
 QY 721 STRAGSEVTIKQEPVSPKKNALRLYLDKDTKDIGLEIPKLERLSDKTPASNTK 780  
 DB 721 STRAGSEVTIKQEPVSPKKNALRLYLDKDTKDIGLEIPKLERLSDKTPASNTK 780  
 QY 781 LIAMKTEKEMSEFEPDQPSLELDNLEIIDLQNSQLPOLFPDTRGAPAVDQOAIL 840  
 DB 781 LIAMKTEKEMSEFEPDQPSLELDNLEIIDLQNSQLPOLFPDTRGAPAVDQOAIL 840  
 QY 841 NDLMQLTAENSPTVPGACQATLRIQSFTNNRPPQGLGILLNQNPLDITTIQSPGAG 900  
 DB 841 NDLMQLTAENSPTVPGACQATLRIQSFTNNRPPQGLGILLNQNPLDITTIQSPGAG 900  
 QY 901 PEPPIRNSPYSVIPOPGMGNGOGMIGNGLNNSSTGMSASRPTEGGEAPQSSA 960  
 DB 901 PEPPIRNSPYSVIPOPGMGNGOGMIGNGLNNSSTGMSASRPTEGGEAPQSSA 960  
 QY 961 VRVTCATTSAMNRPVQGMIRNPPASIPMRPSQPGQRTLOQVYVNTIGPSELENNMG 1020  
 DB 961 VRVTCATTSAMNRPVQGMIRNPPASIPMRPSQPGQRTLOQVYVNTIGPSELENNMG 1020  
 QY 1021 PYSQOQAPPNOTAPPESTLPTIDQASFASQNPRESSFDLLCPRPAESPSDEGALL 1080  
 DB 1021 PYSQOQAPPNOTAPPESTLPTIDQASFASQNPRESSFDLLCPRPAESPSDEGALL 1080

Db 1021 PQYNQQAAPPNQTAFWPESITLPIIDQAFSGSQRHHPFGSSPDDLCLPHPAESPDEGALL 1080  
 Qy 1081 DQLYALANFDLEEDIDRALGIPELVSSQAVDPPEOFSSODSNIMLEOKAPVPEOQVASQ 1140  
 Db 1081 DQLYALANFDLEEDIDRALGIPELVSSQAVDPPEOFSSODSNIMLEOKAPVPEOQVASQ 1140  
 Qy 1141 AQMAAGSYSPMODPNEHTMGORPSYATLTMOPRPLRPTGLVQNPOLRLQJHRLOAQ 1200  
 Db 1141 TOMAGSYNPMDPNEHTMGORPSYATLTMOPRPLRPTGLVQNPOLRLQJHRLOAQ 1200  
 Qy 1201 QNRPLAMQISVSNVNTLRGCVPTQAPINQMLAQORRELLNHLROQHQQOQVQ 1260  
 Db 1201 QNRPLAMQISVSNVNTLRGCVPTQAPINQMLAQORRELLNHLROQHQQOQVQ 1260  
 Qy 1261 RTLMRRGGGLNMTPSMVAAPSGPATMSNPRIQANAAOOPPEPNTGISOOPRGFTGAT 1320  
 Db 1261 RTLMRRGGGLNMTPSMVAAPSGPATMSNPRIQANAAOOPPEPNTGISOOPRGFTGAT 1320  
 Qy 1321 POSPLMSPRMAHTQSPMMQSQANPAYQAPSDINGMAGNMGNSMFSSQSPPHFGQOAN 1380  
 Db 1321 POSPLMSPRMAHTQSPMMQSQANPAYQAPSDINGMAGNMGNSMFSSQSPPHFGQOAN 1380  
 Qy 1381 TSMYSNNNNINVSATNTGMSNNQMTGQISMTSVTSYSTGSLSMGPBYNDPALRG 1440  
 Db 1381 TSMYSNNNNINVSATNTGMSNNQMTGQISMTSVTSYSTGSLSMGPBYNDPALRG 1440  
 Qy 1441 NLF--PNOLPGMDMKOEGDTRKYC 1464  
 Db 1441 SLFTTNOLPGMDMKOEGDTRKYC 1465

RESULT 4  
 NCO2\_XENLA  
 ID NCO2\_XENLA STANDARD: PRT: 1516 AA.

AC Q9W705;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Nuclear receptor coactivator 2 (Transcriptional intermediary factor 2) (XNIF2).  
 GN NCOA2 OR TIF2.  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae; Xenopodidae; Xenopus.  
 OC NCBI\_TaxID=8355;  
 RN [1]  
 RP MEDLINE=20171035; PubMed=10704837;  
 RA de la Calle-Mustienes E., Gomez-Skarmeta J.L.;  
 RT "XNIF2, a Xenopus homologue of the human transcription intermediary factor, is required for a nuclear receptor pathway that also interacts with CBP to suppress Brachyury and Xmyod.";  
 RL Mech. Dev. 91:119-129 (2000).  
 CC -1- FUNCTION: TRANSCRIPTIONAL COACTIVATOR FOR STEROID RECEPTORS AND NUCLEAR RECEPTORS. COACTIVATOR OF THE STEROID BINDING DOMAIN (AF-2) BUT NOT OF THE MODULATING N-TERMINAL DOMAIN (AF-1).  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED HOMOGENEUSLY DURING LATE BLASTULA-EARLY GASTRULA STAGE AND LATER BECOMES HIGHLY EXPRESSED IN THE NOTOCHORD.  
 CC -1- SIMILARITY: CONTAINS 1 PAS (PER-ARNT-SIM) DIMERIZATION DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: AJ243119; CAB45389.1;  
 DR InterPro: IPR001092; HLH\_dlm.

DR InterPro: IPR000014; PAS.  
 DR Pfam: PF00989; PAS; 1.  
 DR SMART: SM00353; HLH; 1.  
 DR SMART: SM00091; PAS; 1.  
 DR PROSITE: PS50112; PAS; 1.  
 KW Transcription regulation; Activator; Nuclear protein.  
 FT DOMAIN 116 180 PAS.  
 FT DOMAIN 1237 1273 POLY-GLN.  
 SQ SEQUENCE 1516 AA; 166156 MW; 09851C00A8439A4A CRC64;

Query Match 69.1%; Score 5276.5; DB 1; Length 1516;  
 Best Local Similarity 69.0%; Pred. No. 8.7e-233;  
 Matches 1058; Conservative 124; Mismatches 225; Indels 127; Gaps 19;

Qy 4 MGEINTSDPSRAETRRKKECPDOLGSPKRNTEKRNREQENKYEELAEILFANFNIDINF 63  
 Db 1 MGEINTSDPSRAETRRKKECPDOLGSPKRNTEKRNREQENKYEELAEILFANFNIDINF 60  
 Qy 64 NFKPKCALTKETVQIOIRIKFEOKAAANIDEVOKSDVSTGGGVYIDKDALGPMYLEAL 123  
 Db 61 NFKPKCALTKETVQIOIRIKFEOKAAANIDEVOKSDVSTGGGVYIDKDALGPMYLEAL 120  
 Qy 124 DGEFFVYNLEGNVNVSENVTOYLRYNOBELANKSVYSILHAGDHTFEYKNILPKSYVNG 183  
 Db 121 DGEFFVYNLEGNVNVSENVTOYLRYNOBELANKSVYSILHAGDHTFEYKNILPKSYVNG 180  
 Qy 184 GSWSGEPPRRNSHTFNCMLVYPLPDSEEGHDNOBAHOKYETMOCFAVSQPKSKEGE 243  
 Db 181 -----VPRRNSHTFNCMLVYPLPDSEEGHDNOBAHOKYETMOCFAVSQPKSKEGE 234  
 Qy 244 DLQSLICVARRVPKERVLPSPSEFTTRDLOGKITSLDSTRBAAMKPGMEDLYRRC 303  
 Db 235 DLQSLICVARRVPKERVLPSPSEFTTRDLOGKITSLDSTRBAAMKPGMEDLYRRC 294  
 Qy 304 IOKFHAOHEGESVYAKRHNEHLROGLAFISOIYRFESLSDGTLVAAOTSKLIRSQTTNE 363  
 Db 295 IOKFHAOHEGESVYAKRHNEHLROGLAFISOIYRFESLSDGTLVAAOTSKLIRSQTTNE 353  
 Qy 364 POLVLSLMLHREOVNVCNPNPLTGTQMGKPLNPISNSPAHQALCSGPNQDMLTSSNI 423  
 Db 354 POLVLSLMLHREOVNVCNPNPLTGTQMGKPLNPISNSPAHQALCSGPNQDMLTSSNI 413  
 Qy 424 NPPINGPKROMKPMRGREGSGGMNHYSGMAQTTPGOSVYALAKMNSPOSSSGMNPQGT 483  
 Db 414 NPPINGPKROMKPMRGREGSGGMNHYSGMAQTTPGOSVYALAKMNSPOSSSGMNPQGT 473  
 Qy 484 SMLSPRHMRSPGVACSPRIPTQSPAGSLHPVGCSSSTGNSHTNSLNAALALSG 543  
 Db 474 SMLSPRHMRSPGVACSPRIPTQSPAGSLHPVGCSSSTGNSHTNSLNAALALSG 533  
 Qy 544 HGVSIGSSIASPDLKMGNLQNSPVNMNPPRLSKMGSLSKDCFGLYGEPSSECTQAESS 603  
 Db 534 HGVSIGSSIASPDLKMGNLQNSPVNMNPPRLSKMGSLSKDCFGLYGEPSSECTQAESS 592  
 Qy 604 CHRGOKETNDNLPRAVSEERADOSRLHDSKGTQLQLLTTSDDQNEPPLAS-SLS 662  
 Db 593 CHRGOKETNDNLPRAVSEERADOSRLHDSKGTQLQLLTTSDDQNEPPLAS-SLS 649  
 Qy 663 DTKNKSTGSLP-----SGSHGTSLKEKHILRLLODSSSPVDLKLTLAEAGKLSQSS 719  
 Db 650 DTKNKSTGSLP-----SGSHGTSLKEKHILRLLODSSSPVDLKLTLAEAGKLSQSS 709  
 Qy 720 SSTAPGEVYTKQEPVSPKKENALLRYLLDKDTRKIDGLPEITPKLERLSKTDPAANT 779  
 Db 710 SSTAPGEVYTKQEPVSPKKENALLRYLLDKDTRKIDGLPEITPKLERLSKTDPAANT 768  
 Qy 780 KLIANKTRKEEMSPRGQPGSELNLTLELIDDLQNSQLPOLFPRTTRGAPAGSDKAI 839  
 Db 769 KLIANKTRKEEMSPRGQPGSELNLTLELIDDLQNSQLPOLFPRTTRGAPAGSDKAI 826  
 Qy 840 INDLQQLTAENSPVPVPGAK-TALRISQST-FNNPRPGQLRLPNQNLPLDITLQSPRT 897  
 Db 839 INDLQQLTAENSPVPVPGAK-TALRISQST-FNNPRPGQLRLPNQNLPLDITLQSPRT 826

Db 827 MNDLMOLAGENSTGLPAPMAOKORMLRMQONNGNSOLAOLGR-LPNOULPLDIFHQSOA 885  
 QY 898 GAGPPPIRNSSPSYVPOGMMNGOMIGNOGNLGNSGTGMIGNSASRPTMPSGEWAPQ 957  
 Db 886 SASPSFAMRSSRYTYVPGSGVINQAMMSQGNVRNRSRGITYGVNPRRPLRPGDMGSO 945  
 QY 958 SSAVRVTCATTSAMNRPVGGMIRNPASIPMRSSQPGOROTLOSQVNNIGPSELENN 1017  
 Db 946 ASAVRPACPTSTAMNR---HDMRSRPTASIPMRPGSQVCPROYLSAVANNMSSSELDNM 1002  
 QY 1018 MGSPQYSQQAAPPNQTAPWPESTILPDAQSAFQONRQPESSSDLLCPHRAESPDG 1077  
 Db 1003 ISGPQYTTQQAAPPNQTAPWPNRLLTLEQPSFNQONRQPESSPADLLCPITVSEPADG 1062  
 QY 1078 ALLDQYTLARNPDGLLEIRALGIPELVQSQAADPEQSSODSNIMLQKAPVFFQOY 1137  
 Db 1063 NLLDQYTLARNPDGLLEIRALGIPELVQSQAADPEQSSODSNIMLQKAPVFFQOY 1122  
 QY 1138 ASQAQMAQSGYSFPMQDPEFTMGORPSYATLRMQPRPGLRPTGLVQNPQNLRLQLOHRL 1197  
 Db 1123 ANQOMQANQSYQPMQDPEFTMGORPSYATLRMQPRPGLRPTGLVQNPQNLRLQLOHRL 1182  
 QY 1198 QAQONRQPLMNOISNYSNNVLTLPGRP---TQAPINAQMLAQORELLINQHLRQROM 1253  
 Db 1183 QA-QNRQQLMNPINNVSNNMLAMPVGPGLRQGPINMQORORELLISQHLRQLOLO 1241  
 QY 1254 QQQQVQO-----RTLMRQGLINMPTSNVAPSGMPATMS 1287  
 Db 1242 QQQ 1301  
 QY 1288 NRPTRQANAOQFFPPPT-----RTLMRQGLINMPTSNVAPSGMPATMS 1305  
 Db 1302 SPRPQSTQOQFFPPPTNYGTGIPSPPTSPFSPVPPSPQSLSHSLGSGQMLNAG 1361  
 QY 1306 -----GISQPPDPTGATTPOSPPLMSPRMAHOTOSPM 1338  
 Db 1362 IMSGMGQYPPVNPQMHNAFQFANSQMSQSDPCTGATTPOSPPLMSPRMAHOTOSPM 1421  
 QY 1339 QOSQANPAYQAPSDINGMAQGNMGNSMFQSPHFGQOANTSMY-SNNMNINVSMTN 1397  
 Db 1422 QOSQANPAYQ-SELNMAQGNPAGNSMFQSPHFGQOANTSMY-SNNMNINVSMTN 1479  
 QY 1398 TCGSSNMONTGQISMTSVTSVTSGLSGMGPED 1431  
 Db 1480 GNGNNNNMONTGQISMTSVTSVTSGLSGMGPED 1513  
 RESULT 5  
 SIMA\_DROME STANDARD; PRT; 1507 AA.  
 ID Q24167; Q9VNA5;  
 AC 15-DEC-1998 (Rel. 37, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Similar protein.  
 GN SIMA OR CG7951.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachytera; Muscomorpha;  
 OC Phyloroidae; Drosophilidae; Drosophila.  
 OC NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96269413; Pubmed=8682312;  
 RA Nambu J.R., Chen W., Hu S., Crews S.T.;  
 RT "The Drosophila melanogaster similar bHLH-PAS gene encodes a protein  
 RT related to human hypoxia-inducible factor 1 alpha and Drosophila  
 RT single-minded.";  
 RL Gene 172:249-254(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; Pubmed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Branton R.C., Rogers J.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Plankoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Correll J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mallet B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Mostrelti A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Munz D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weisenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 CC -1- FUNCTION: POSSIBLE DNA-BINDING TRANSCRIPTIONAL ACTIVATOR.  
 CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER  
 CC BHLH PROTEIN.  
 CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).  
 CC -1- TISSUE SPECIFICITY: UBICITOUSLY EXPRESSED IN THE EMBRYO.  
 CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF  
 CC TRANSCRIPTION FACTORS. HIGHEST TO HIF-1 ALPHA.  
 CC -1- SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U43090; AAC47303.1; -  
 DR EMBL: AE003772; AAF57008.2; -  
 DR Flybase: FBgn0015542; sima.  
 DR InterPro: IPR003015; HLH\_Myc.  
 DR InterPro: IPR001092; HLH\_dim.  
 DR InterPro: IPR001610; PAS.  
 DR InterPro: IPR000014; PAC.  
 DR Pfam: PF00785; PAC; 1.  
 DR Pfam: PF00989; PAS; 2.  
 DR SMART: SM00053; HLH; 1.  
 DR SMART: SM00086; PAC; 1.  
 DR SMART: SM00091; PAS; 2.  
 DR PROSITE: PS00038; HELIX\_LOOP\_HELIX; 1.  
 DR PROSITE: PS00112; PAS; 2.  
 KW Repeat; DNA-binding; Nuclear protein; Transcription regulation;  
 KW Activator; Coiled coil.  
 FT DNA\_BIND 72 85 BASIC DOMAIN.



```

OY 1339 -----OOSQANPAYQAPSDINGMAQNMGSFMSQOSPPHFGQANTSMYNNMNIN 1391
DB 1247 QLICKPAAQOQOAKRASE--RW-----QLSAESK-----QOKQOQOQSNVLAN 1291
OY 1392 VSMATNTGMSMSNMONTGOISMTSVTSVSTGL-----SSMGPEQVNDPALRGNEF 1443
DB 1292 LVSQGRDDDDSEAMTIDEDNSLVQPIPLGKYGPLCHTSTSVLRDYNHNNPLISGTNF- 1350
OY 1444 PNDLP-----GMDKROEGDT 1459
DB 1351 --QLSPVFGSDSSGGDGET 1368

RESULT 6
CLOC_HUMAN STANDARD; PRT; 846 AA.
AC O1516; O14516; Q9UTR8;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Circadian locomotor output cycles kaput protein (hCLOCK).
GN CLOCK OR KIAA0334.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA King D.P., Steeves T.D.L., Zhao Y., Sangoram A.M., Takahashi J.S.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA MEDLINE=99216412; PubMed=10198158;
RA Steeves T.D.L., King D.P., Zhao Y., Sangoram A.M., Du F.,
RA Bowcock A.M., Moore R.Y., Takahashi J.S.;
RT Molecular cloning and characterization of the human CLOCK gene:
RT expression in the suprachiasmatic nuclei.";
RL Genomics 57:189-200(1999).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=97349984; PubMed=9205841;
RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. VII.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 4:141-150(1997).
RN [4]
RP SEQUENCE OF 1-349 FROM N.A.
RC TISSUE=Brain;
RA Ikeda M., Takehara N., Edisawa T., Yamauchi T., Nomura M.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
CC - FUNCTION: CIRCADIAN REGULATOR THAT ACTS AS A TRANSCRIPTION FACTOR.
CC CLOCK-BMAL1 HETERODIMERS BIND TO AN E-BOX ELEMENT (3'-CAGGTG-5'),
CC THEREBY ACTIVATING TRANSCRIPTION OF PER1, AND POSSIBLY OF OTHER
CC CIRCADIAN CLOCK PROTEINS. MUTANT CLOCK AND BMAL1 FORM HETERODIMER
CC THAT BIND DNA, BUT FAIL TO ACTIVATE TRANSCRIPTION (BY SIMILARITY).
CC - SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER
CC BMH1 PROTEIN. HETERODIMER WITH BMAL1, AND LESS EFFICIENTLY WITH
CC ARNT AND ARNT2. HETERODIMERS WITH ARNT OR ARNT2 BIND POORLY TO THE
CC E-BOX MOTIF (BY SIMILARITY).
CC - SUBCELLULAR LOCATION: Nuclear (Potential).
CC - SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS.
CC - SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.
CC - SIMILARITY: CONTAINS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: AF011568; AAB83969.1; -.
DR EMBL: AF097458; AAF13733.1; -.
DR EMBL: AF097442; AAF13733.1; JOINED.
DR EMBL: AF097443; AAF13733.1; JOINED.
DR EMBL: AF097444; AAF13733.1; JOINED.
DR EMBL: AF097445; AAF13733.1; JOINED.
DR EMBL: AF097446; AAF13733.1; JOINED.
DR EMBL: AF097447; AAF13733.1; JOINED.
DR EMBL: AF097448; AAF13733.1; JOINED.
DR EMBL: AF097449; AAF13733.1; JOINED.
DR EMBL: AF097450; AAF13733.1; JOINED.
DR EMBL: AF097451; AAF13733.1; JOINED.
DR EMBL: AF097452; AAF13733.1; JOINED.
DR EMBL: AF097453; AAF13733.1; JOINED.
DR EMBL: AF097454; AAF13733.1; JOINED.
DR EMBL: AF097455; AAF13733.1; JOINED.
DR EMBL: AF097456; AAF13733.1; JOINED.
DR EMBL: AB002332; BAA20792.1; -.
DR EMBL: AB005335; BAA21774.1; -.
DR HSSP: P36956; IAW9.
DR MIM: 601851; -.
DR InterPro: IPR003015; HLH_MYC.
DR InterPro: IPR001092; HLH_dlm.
DR InterPro: IPR001067; Nucleinslocator.
DR InterPro: IPR001610; PAC.
DR InterPro: IPR000014; PAS.
DR Pfam: PF00785; PAC; 1.
DR Pfam: PF00989; PAS; 2.
DR PRINTS: PR00785; NCTRNSLOCATR.
DR SMART: SM00353; HLH; 1.
DR SMART: SM00086; PAC; 1.
DR SMART: SM00091; PAS; 2.
DR PROSITE: PS00038; HELIX_LOOP_HELIX; 1.
DR PROSITE: PS0112; PAS; 2.
DR KW Transcription regulation; Nuclear protein; Repeat; Biological rhythms;
KW DNA-binding.
FT DNA_BIND 35 47 BASIC DOMAIN.
FT DOMAIN 48 85 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
FT DOMAIN 107 177 PAS 1.
FT DOMAIN 262 332 PAS 2.
FT DOMAIN 514 564 IMPLICATED IN THE CIRCADIAN RHYTHMICITY
(BY SIMILARITY).
FT DOMAIN 744 760 GUN-RICH.
FT DOMAIN 819 828 POLY-GLN.
FT CONFLICT 440 440 S -> P (IN REF. 2).
SQ SEQUENCE 846 AA; 95303 MW; C292B451A33EACBF CRC64;

Query Match 4.2%; Score 317; DB 1; Length 846;
Best local Similarity 21.5%; Pred. No. 1.5e-07;
Matches 176; Conservative 129; Mismatches 316; Indels 196; Gaps 39;

OY 31 KRTEKRNQEQKYTEELAELEIFANFNIDNNFNPKDPKCALIKETVKOIRQKDEKAA 90
DB 39 RKNSEKRRDQEQVLLKELGSMPLGNAR-----KMDKSTYLQKSIDPLR-KHKEITR 89
OY 91 AANIDQKSDVSTSGGVYDKDALGPMMLDEALDGFPPVNVLEGNVVFSENYTOYLRYN 150
DB 90 QSDASRL-RQDKWPT--PLSNELFQJMEALDGFFLAIMDGSIIYSEVSTLSLEHL 145
OY 151 QEELMNKSVYSILVGDHTEFYVKNLLPKSTVNGSVSGEPPR-RNSHTYNCRLM----- 203
DB 146 PSDLVQDSITNFPEGEHSEVYK-ILSTHLESDSLTPREYLSKKNOLFECCHMLRGRTIDP 204
OY 204 -----VRLPDSSEEGHDN-----QEAHQ-KYELMOQFA-----VSQKSTK 239
DB 205 KEPSTVEYVKIGNFKSLNSVSSAHNGFEGTIGRTHRPSYEDRVGCVATVRLATPOFK 264

```









Cell 89:641-653(1997).

[3] SEQUENCE FROM N.A. STRAIN129/SV; Wilsbacher L.D., Sangoram A.M., Antoch M.P., Takahashi J.S.; "The mouse clock locus: Sequence and analysis of 204 kb from mouse chromosome 5"; Submitted (May 2000) to the EMBL/GenBank/DBJ databases.

[4] INTERACTION WITH BMAL1. MEDLINE=98279137; Pubmed=9616112; Gekakis N., Staknis D., Nguyen H.B., Davis F.C., Wilsbacher L.D., King D.P., Takahashi J.S., Weitz C.J.; "Role of the clock protein in the mammalian circadian mechanism."; Science 280:1564-1569(1998).

-1- FUNCTION: CIRCADIAN REGULATOR THAT ACTS AS A TRANSCRIPTION FACTOR. CLOCK-BMAL1 HETERODIMERS BIND TO AN E-BOX ELEMENT (3'-CACGTG-5'), THEREBY ACTIVATING TRANSCRIPTION OF PER1, AND POSSIBLY OF OTHER CIRCADIAN CLOCK PROTEINS. MUTANT CLOCK AND BMAL1 FORM HETERODIMER THAT BIND DNA, BUT FAIL TO ACTIVATE TRANSCRIPTION. IN HOMOLOGOUS CLOCK MUTANTS, THE CIRCADIAN PERIOD IS INCREASED FROM 3 TO 4 HOURS AND USUALLY THE CIRCADIAN RHYTHMICITY IS LOST AT CONSTANT DARKNESS. EXPRESSION OF CLOCK IS ALSO REDUCED.

-1- SUBUNIT: HETERODIMER WITH BMAL1, AND LESS EFFICIENTLY WITH ARNT AND ARNT2. HETERODIMERS WITH ARNT OR ARNT2 BIND POORLY TO THE E-BOX MOTIF.

-1- TISSUE SPECIFICITY: EXPRESSED EQUALLY IN BRAIN, EYE, TESTES, OVARIES, LIVER, HEART, LUNG, KIDNEY. IN THE BRAIN, EXPRESSION IS ABUNDANT IN THE SUPRACHIASMATIC NUCLEI (SCN), IN THE PYRIFORM CORTEX, AND IN THE HIPPOCAMPUS. LOW EXPRESSION THROUGHOUT THE REST OF THE BRAIN. EXPRESSION DOES NOT APPEAR TO UNDERGO CIRCADIAN OSCILLATIONS.

-1- DOMAIN: CONTAINS A GLN-RICH C-TERMINAL DOMAIN WHICH COULD CORRESPOND TO THE TRANSACTIVATION DOMAIN. IN MUTANT CLOCK, DELETION OF THIS REGION LEADS TO AN INCREASED CIRCADIAN PERIOD FROM 3 TO 4 HOURS AS WELL AS TO THE LOSS OF CIRCADIAN RHYTHMICITY AND ALTERED LIGHT RESPONSE.

-1- DISEASE: DEFECTS IN CLOCK AFFECT TWO PROPERTIES OF THE CIRCADIAN SYSTEM: THE LENGTH OF THE FREE-RUNNING PERIOD AND THE PERSISTENCE OF CIRCADIAN RHYTHMICITY IN CONSTANT DARKNESS.

-1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF TRANSCRIPTION FACTORS.

-1- SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.

-1- SIMILARITY: CONTAINS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.

-----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation in the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL: AF004998; AAC53200.1; -  
EMBL: AF146793; AAC30565.1; -  
MGD: MGI:99698; CLOCK.

DR InterPro: IPR003015; HLH\_Myc.  
DR InterPro: IPR001092; HLH\_dim.  
DR InterPro: IPR001067; NucleusLocator.  
DR InterPro: IPR001610; PAC.  
DR InterPro: IPR000014; PAS.  
DR InterPro: IPR003617; TFS2\_N.  
DR Pfam: PF00785; PAC; 1.  
DR Pfam: PF00989; PAS; 2.  
DR PRINTS: PRO0785; NCMRNSLOCATR.  
DR SMART: SM00353; HLH; 1.  
DR SMART: SM00086; PAC; 1.  
DR SMART: SM00091; PAS; 2.  
DR SMART: SM00509; TFS2N; 1.  
DR PROSITE: PS00038; HELIX\_LOOP\_HELIX; 1.  
DR PROSITE: PS50112; PAS; 2.  
DR Transcription regulation; Nuclear protein; Repeat; Biological rhythms; DNA-binding; Alternative splicing.

Query Match	Best Local Similarity	Conservative	Score	DB	Length
Matches 150;	21.7%;	107;	309.5;	181;	85;
31	KRNTKRRKREDENKITEELAEIPANFNFDIDNFNFKPRKCAIKETVQIROIKOECAA	90			
39	RKRSKRRKRDQFNVLKELGSLPENAR---KMKSTVLOKSIDFLR--KIKETETA	89			
91	AANIDVOKSDVSSSGGVINDKDALGAPMALDGEFFVNVLEGVNVVSENVTOYLRYN	150			
90	OSDASEI--KODMKP---FLSNEETQJMLERLDGFFLAIMTDSIITYVESVYSLEHL	145			
151	QELINKSVYSILHYHDHTEFYKMLPKRSIYNGSGWSGSEPPR--RNSHTFNGRML-----	203			
146	PSDLVDGSIFFNIPRGEHSEYVK--ILSHLHLESDSLPEYLKSKNQLEFCCHMLGTDIP	204			
204	-----VKRPLPDEEGBHDN-----QEAH-KIETMOCFA---VSOPKSIK	239			
205	KEPSTVEYVFIRGNKSLTYSVSTHNGFEQIORTHRPSYEDRVCFAVATVRLAPOPFIK	264			
240	EEGEDJGSLICVAVARVPMKERPLVPSSESTTRDGLKITSIDTSTMRAKRGWEDL	299			
265	E-----MCTV-----EER-----NEETTSRLSLWKFLFLD--HRAPIITGIYLF	302			
300	VRRCIOKTHAHEGSESVARRRHHNEVLDRGLAFSOIARFSLSDGTILVAQTKSLIRSQ	359			
303	EVLGTSGYDYHVDLDLMLACHEH--LMQYGGKSCYYRPLTKGGQWIMQTHYITYHQ	361			
360	TTNEQOLVYISLMLHREDOVCVAMPDLTGQTMGKPLNFSSNSPAHQALCSGNPGQDMFL	419			
362	WNSREFIVCTHTVVSVAEVAERREELG-----IEESLPETADKSDSGSDMRI	412			
420	SSNINFPINGRKEQMGPMGRGSGGANNHYSGMATTPOGSNVALKKNSS--PSQSPGMN	478			
413	NT-----VSLKALELRDHS-----PPSASSRSKRSKSHTPNADSPST	451			
479	PGO--PTSMLSP-----RHRMSPGVASPRIPSPQSPAGSLSPVGVCSST	523			
452	PTKIPTDTSTPRQHLPAHEKMTQGRSSFFSSQINSQSVGRSLTPQPAHQANLPI--PQ	509			
524	GNSHSYTSSSLNALQALSE-----GHGVSGLSSL	552			
510	GMSQFOFSAOLGAMQHLKDLEORTRMIEANIHRQEEELRRKIQEOLQWVHQGLQDMFLQ	569			
553	ASPDLMKMNILQNS-----PVNNMPP--PLSKMSGLDSKOCFGILGERSBS-----	595			
570	SNPGINFSVYDLSGNSNIQOLTPVNMGGQVYPAQVOS-----GHISTGQHMIIQ	619			
596	-----TTGOAESCHPGEQETNDPMLPPA	620			
620	QQTLOSTSTQSQSQSVMSGHSQQTSLPSQTS	651			
9	NP2A MOUSE	STANDARD;	PRT;	816 AA.	
15-DEC-1998 (Rel. 37, Created)					
15-DEC-1998 (Rel. 37, Last sequence update)					

DT 16-OCT-2001 (Rel. 40, last annotation update)  
DE Neuronal PAS domain protein 2 (Neuronal PAS2).  
GN NPAS2.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
ON NCBI\_TaxID=10090;  
RX (1)  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RA MEDLINE=97165088; PubMed=9012850;  
RA Zhou Y.-D., Bernard M., Tian H., Li X., Ring H.Z., Francke U.,  
RA Shelton J., Richardson J., Russell D.W., McKnight S.L.;  
RT "Molecular characterization of two mammalian bHLH-PAS domain proteins  
RT selectively expressed in the central nervous system.";  
RL Proc. Natl. Acad. Sci. U.S.A. 94:713-718(1997).  
CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER  
CC bHLH PROTEIN. INTERACTS WITH HSP90.  
CC -1- SUBCELLULAR LOCATION: Nuclear (Potentially).  
CC -1- TISSUE SPECIFICITY: IN BRAIN, EXCLUSIVELY NEURONAL. ALSO FOUND IN  
CC SPINAL CORD, AND IN A LESSER EXTENT IN COLON, SMALL INTESTINE AND  
CC UTERUS.  
CC -1- DEVELOPMENTAL STAGE: FIRST DETECTED 3 DAYS AFTER BIRTH.  
CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (bHLH) FAMILY OF  
CC TRANSCRIPTION FACTORS.  
CC -1- SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.  
CC -1- SIMILARITY: CONTAINS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL: U77969; AAB47249.1; -.  
DR MGD; MG1:109232; NPas2.  
DR InterPro: IPR003015; HLH\_Myc.  
DR InterPro: IPR001092; HLH\_dlm.  
DR InterPro: IPR001610; PAC.  
DR InterPro: IPR000014; PAS.  
DR Pfam: PF00010; HLH; 1.  
DR Pfam: PF00785; PAC; 1.  
DR Pfam: PF00989; PAS; 2.  
DR SMART: SM00353; HLH; 1.  
DR SMART: SM00086; PAC; 1.  
DR SMART: SM00091; PAS; 2.  
DR PROSITE: PS00038; HELIX\_LOOP\_HELIX; 1.  
DR PROSITE: PS50112; PAS; 1.  
KW Repeat; DNA-binding; Nuclear protein; Transcription regulation.  
FT DNA\_BIND 10 22 BASIC DOMAIN.  
FT DOMAIN 23 60 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).  
FT DOMAIN 82 152 PAS 1.  
FT DOMAIN 237 307 PAS 2.  
FT DOMAIN 311 354 PAC.  
SQ SEQUENCE 816 AA; 90915 MW; 7E5CF0641CFDC1DD CRC64;  
  
Query Match 3.8%; Score 289; DB 1; Length 816;  
Best Local Similarity 18.3%; Pred. No. 2, 7e-06;  
Matches 211; Conservative 151; Mismatches 360; Indels 432; Gaps 42;

Db 94 DGFVIVTTDGSIIIVSDITPLLGHLPADVDMDNLLNPLPEQSEHYVKKIISNHLVTD 153  
Qy 184 GSWSCGEPFRNRSHTFNCRLV---VKP-----LPDSEEGHND--- 217  
Db 154 SPSEPEFLKSDNDELFCYCHLLRSLNKEPTTYEIKFVGNFRYNNVPSPSCGFPTNTS 213  
Qy 218 QEAHQKYTEWQCF---AVSQPKIEEGEDLOSLICVARRVPKMERPLVPSSEFTTR 273  
Db 214 RQCHVPLGKVCFCIATVRLATPQFLKE-----MCVA-----DEPL---EEFTSR 254  
Qy 274 QDQGAITSLDITSTMRAANKPGW---EDLYRCIQKFEHQHEGESVYAKRHNNHEVLRG 330  
Db 255 HSLWKFFLELD---HRAPIIGLPEVEVLGTSGYNYTH---DLELARQHNLHQFG 307  
Qy 331 LAFSQIYRESLSDGTVAQAQTSKILRSQTNPEQLVISLH-----MLHREQVCMY 382  
Db 308 KKGSCCYRFLTKQOMIWIQTHYIITYHQMNSKEPIVCHSVSYADRVREKQDLALE 367  
Qy 383 NPDLTGQTMGKPLNPISNSSPAHQALCSGNPGQDMLSSNINFPINGPKQMGMPGRG 442  
Db 368 DP-----PTEAM----- 374  
Qy 443 GSGGMHVSGMQATTPQGSNYALKMNSPQSSPGMPCOPTSMLSPRHMSPGVAGSPRI 502  
Db 375 -----HPSAVK-----EKDSLPEPQFNL-----DMGASLPPSS 405  
Qy 503 PPSQSPAGSLHSPVGCSTGNSHSTYNSLALQALSGHGVSLGSSIASPLDKMGNL 562  
Db 406 P-----SPSASSRS-----SHKSHTRAMS----- 424  
Qy 563 QNSPVMNPPPLSKMGSLSKDCFGLYGER-SEGTTCQAESSCHPEQKETNDPNLPVAV 621  
Db 425 -----EPTSPPTLMENSTYALPRATLPEQLPELVQ 456  
Qy 622 SSERADQGSRLHDSKG---QTKLQLLTTKSDMPSPPLAS-----SLSDTKDSTGSL 672  
Db 457 LSOAATMPALHSSASCDLTQKLLQSLPOTGLQSPAPVPTQPSAOFSMQYTKD----- 511  
Qy 673 PGSGSTHGTSLEKHKILHRLLDSSSPVDLAKLTAEATGKDLSSSTAPSGEYTIQ 732  
Db 512 -----QLEGRTRLQ-----ANIRMG 528  
Qy 733 EPVSPKKENALRYLLDKDQTDIGLPEITPKLERLDSKTPASNTKLIAMTEKEENS 792  
Db 529 EELHAKIQEQLCIQV-----DSNVQMF-LQQPAVSLSS 558  
Qy 793 FEEGDDPGSELNLELIDLDQNSQLPQLPDRPGAPAGSVKQALINDLMQLTANSP 852  
Db 559 FSTQRPAAQ-----QQLQQ-----QPLAVN----- 586  
Qy 853 VTPVGAQKTALRISQS-----TFNNRPQGLRLLPNQMLPL-DITLQSGPTGAGPF 902  
Db 587 -TFLQGLITSTQYTNGLHRESNVISAQPKPARSSQLLPASGRSLSPSGSYSAVYL 645  
Qy 903 PIRNSSPYSVIPQPGMNGMIGNOGNIENS-STGMIGNSASRPTMPSGEAPQSSAV 961  
Db 646 PQLISLTTIATPQ-----DSSOCPSPDGHRQLRLILSQIOPMPPSSCARPSEV 700  
Qy 962 RVT-----CAATSAMNRPVCGMIRNPAASIRPPSSOPGQRTLDS 1004  
Db 701 SKRGROYKAYQASQVMPPSPSHPTNSASAPV---LLMGQAVLHPSFSPSRPPLD----- 753  
Qy 1005 QVANNIGPSELEMMNGGPOYSQQQAAPPNQAP-----WPESILPIQASFAASQNRQPFGS 1058  
Db 754 -----PAQAQQQPPPYLQAPSLHSEBQPSLL-----LSFTSQQPTQITG 793  
Qy 1059 SPDDLCLPHPAADS 1072  
Db 794 AATQSTPPQPPRS 807  
  
RESULT 10  
NFT5\_HUMAN

ID NPT5\_HUMAN STANDARD; PRT; 1531 AA.  
AC O94916; O95693; O9UN18;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Nuclear factor of activated T cells 5 (T cell transcription factor  
DE NFAT5) (NF-AT5) (Toxicity-responsive enhancer-binding protein) (Tone-  
DE binding protein) (ToneBP).  
GN NFAT5 OR TONEBP OR KIA0827.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_taxid:9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM C).  
RC TISSUE-Brain;  
RX MEDLINE-99156230; PubMed-10048485;  
RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirosewa M.,  
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;  
RT "Prediction of the coding sequences of unidentified human genes. XII.  
RT The complete sequences of 100 new cDNA clones from brain which code  
RT for large proteins in vitro.";  
RL DNA Res. 5:355-364(1998).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM A).  
RC TISSUE-Brain;  
RX MEDLINE-99307389; PubMed-10377394;  
RA Lopez-Rodriguez C., Aramburu J., Rakeam A.S., Rao A.;  
RT "NFAT, a constitutively nuclear NFAT protein that does not cooperate  
RT with Fos and Jun.";  
RL Proc. Natl. Acad. Sci. U.S.A. 96:7214-7219(1999).  
RN [3]  
RP SEQUENCE FROM N.A. (ISOFORM A).  
RX MEDLINE-99162641; PubMed-10051678;  
RA Miyakawa H., Woo S.K., Dahl S.C., Handler J.S., Kwon H.M.;  
RT "Tonically-responsive enhancer binding protein, a rel-like protein that  
RT stimulates transcription in response to hypertonicity.";  
RL Proc. Natl. Acad. Sci. U.S.A. 96:2538-2542(1999).  
RN [4]  
RP SEQUENCE FROM N.A. (ISOFORM B).  
RC TISSUE-Brain;  
RA Lopez-Rodriguez C., Aramburu J., Rakeam A.S., Copeland N.G.,  
RA Gilbert D.J., Thomas S., Distche C., Jenkins N.A., Rao A.;  
RT "NFAT5: The NF-AT family of transcription factors expands in a new  
RT direction.";  
RL Cold Spring Harb. Symp. Quant. Biol. 64:517-526(1999).  
RN [5]  
RP SEQUENCE OF 675-1531 FROM N.A.  
RC TISSUE-Brain;  
RX MEDLINE-20029268; PubMed-10565538;  
RA Zuehlke C., Kiehl R., Johannsmeyer A., Grzeschik K.H., Schwinger E.;  
RT "Isolation and characterization of novel CAG repeat containing genes  
RT expressed in human brain.";  
RL DNA Seq. 10:1-6(1999).  
CC -1- FUNCTION: PLAYS A ROLE IN THE INDUCIBLE EXPRESSION OF GENES.  
CC -1- REGULATES HYPERTONICITY-INDUCED CELLULAR ACCUMULATION OF  
CC OSMOLYTES.  
CC -1- SUBUNIT: DOES NOT BIND WITH FOS AND JUN TRANSCRIPTION FACTORS. BUT  
CC MIGHT BE CAPABLE OF FORMING STABLE DIMERS WITH DNA ELEMENTS.  
CC -1- SUBCELLULAR LOCATION: Nuclear.  
CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS: A, B AND C (SHOWN HERE); MAY BE  
CC PRODUCED BY ALTERNATIVE SPLICING.  
CC -1- TISSUE SPECIFICITY: HIGHEST LEVELS IN SKELETAL MUSCLE, BRAIN,  
CC HEART AND PERIPHERAL BLOOD LEUKOCYTES. ALSO EXPRESSED IN PLACENTA,  
CC LUNG, LIVER, KIDNEY, PANCREAS, SPLEEN, THYMUS, PROSTATE, TESTIS,  
CC OVARY, SMALL INTESTINE AND COLON.  
CC -1- SIMILARITY: BELONGS TO THE REL/DORSAL FAMILY.  
CC -1- CAUTION: REF.5 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 1165  
CC ONWARD DUE TO A FRAMESHIFT.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
CC EMBL; AB020634; BAA74850.1; -  
CC EMBL; AF089824; AAD18136.1; -  
CC EMBL; AF134870; AAD38360.1; -  
CC EMBL; 297016; CA009693.1; ALT\_FRAME.  
CC EMBL; AF163836; AAD48441.1; -  
CC MIM; 604708; -  
CC InterPro; IPR002909; IPT\_TIG.  
CC InterPro; IPR000451; IPT.  
CC Pfam; PF01833; TIG; 1.  
CC SMART; SM00429; IPT; 1.  
CC PROSITE; PS01204; REL\_1; FALSE\_NEG.  
CC PROSITE; PS0254; REL\_2; 1.  
CC Transcription regulation; Activator; Nuclear protein; DNA-binding;  
CC Alternative splicing.  
CC DOMAIN 69 100 SER-RICH. DNA-BINDING.  
CC DOMAIN 293 300 POLY-GLN.  
CC DOMAIN 739 743 POLY-GLN.  
CC DOMAIN 879 888 POLY-GLN.  
CC DOMAIN 966 971 POLY-THR.  
CC DOMAIN 1248 1266 POLY-GLN.  
CC VARSPLIC 1 76 MISSING (IN ISOFORM A).  
CC CONFLICT 1369 1369 E -> D (IN REF. 5).  
CC SEQUENCE 1531 AA; 165764 MW; A68C6808BDBAF69E CRC64;  
  
Query Match 3.8%; Score 289; DB 1; Length 1531;  
Best Local Similarity 20.9%; Pred. No. 6,1e-06;  
Matches 292; Conservative 164; Mismatches 545; Indels 396; Gaps 68;  
  
OY 219 EAHQKYEIMOCFAVQSPQPSIKKEGEDLQSLICVARRVPMKRPVLPSESEFTTQDLOG 278  
DB 319 ECHNEPVVLQVGVGDSGRVKKPHG-FYQACRYTGANTTPCKE-----VDIEG 364  
  
OY 279 KITSIDTSTMRAMKPGME-DLVRRICQFNAHGESESVYAKRHNEV-LRQGLAFSQ- 335  
DB 365 -----TTVIEVGIDRPSNNMGLAVDC-----VGILKLNADVEARIGIAGSK 406  
  
OY 336 -----IYRPSL--SDGTIVAQR-KSLIRSOFTNEPOLY-LSLH--MLHREONCYM 382  
DB 407 KSTRALVFRVIMKRKDSFTLTQTPSSPTLCTOPAGVPEILKSLHSCSVAGEEVR-- 464  
  
OY 383 NPDLTGOTMGRPLNPDISSNSPAHQALCSGNPGQDWTLSNIFPINC---KEQGMIPMG 439  
DB 465 ---LIGKMFLOKTKYIPQENSDENSKMSSEALIDMELFHQNLILYKVPYRHQHTLTVS 521  
  
OY 440 ---RFGSGGNNHVSQMATTPOGSNYALKMNSPQSSPCGNPQPTSMLSPRHMS-- 493  
DB 522 VGIVVVTNAGRSHDQPTTPYPPDPAAGALNVVKKIS---SPARCSFEEBAMKMTT 578  
  
OY 494 -----PGVAGSPHLPSPQSPASLHSPVCVCSGTGSHY--INSSLMALQALSE 542  
DB 579 GGNLDKVNITPVALMTPPLIPSSMIKSEDV--TPMEVTAERKSTTFKTKSVGSYQDTLE 636  
  
OY 543 -----GHG-VLSGSSLASPLDKMGNLQNSPVNMPPLSKGSLSDSCFELYGEPESEG 595  
DB 637 NISNLAGNGSPSPSSHLPSSENEKQQLQPKAYNPEL----- 675  
  
OY 596 TTGAQESSCHRG-----EOKETNDPNLPPAVSSERADGOSR-LNDSKQOTKLLOLT 646  
DB 676 TTIQTODISQCTFPFPAVSASSQLPNSDALLQATQFQTRQTSREILQSDGVVNLISLT 735  
  
OY 647 TKSDQMEPSPL-----ASSLSQTNKD-STGSLPGSGSTGTGTLK 684  
DB 736 EASQOQOOSPLQEQACTLQOQISSNIPSPNSVSQDQNTIQQLOLQSGSTGTGAS----- 789  
OY 685 EKXKILHLRLDSSSPVDLAKLTAA-----TGKDLQSSSSTPAGSEVTTIKQ 732



```
FT CONFLICT 27 27 D -> G (IN REF. 1).
FT CONFLICT 51 51 V -> M (IN REF. 1).
FT CONFLICT 349 349 O -> S (IN REF. 2).
FT CONFLICT 368 368 G -> GG (IN REF. 3).
FT CONFLICT 374 374 P -> S (IN REF. 2).
FT CONFLICT 393 393 P -> L (IN REF. 2).
FT CONFLICT 489 502 MAMPPOVSGMCP -> NGQYATSGWVDVS (IN REF. 2).
FT CONFLICT 634 634 E -> G (IN REF. 2).
FT CONFLICT 924 924 P -> S (IN REF. 2).
FT CONFLICT 1016 1016 Q -> L (IN REF. 2).
FT CONFLICT 1033 1033 Q -> P (IN REF. 2).
SQ SEQUENCE 1902 AA: 205946 MW: 9870A46F81062EAA CRC64;
```

Query Match 3.7%; Score 284.5; DB 1: Length 1902;  
Best Local Similarity 21.4%; Pred. No. 1.3e-05;  
Matches 289; Conservative 126; Mismatches 450; Indels 483; Gaps 74;

```
OY 205 KPLPD--SEEGHNDQEAHOKET---MQCF-----AVSQPKSKEEG--E 243
Db 184 QPASTLSQAAYRPOQSOQOTAYSOQRPPPELSODSGSAASPMTSSKGGE 243
OY 244 DLQSLICVARRVPMKERVLPSSSEFTTRDQKITSLDTSTNRAMKPGMEDLVRC 303
Db 244 DM-----NLSQSNR-----SSLPLDGSIDDLPMGT-EGALSPG----- 277
OY 304 IQKHAHGEHSVYAKHHHEVLRQGLAFSOIYRFSLSDGTLVAQTKSLIRSQTNE 363
Db 278 -----VSTSGISSSQ-----GEQSNP 293
OY 364 POLVLSLMLHREQNVCMNPDLTGTMGKPLNPISNSPAAHQALCSGPGQDMTSSNI 423
Db 294 AOSPPSPH-----TSPHLPG-INGPSPSPV--GSPASVAAQSRGCP-----LS--- 332
OY 424 NFPIPGKQGMKMPKRGPGSGGNHVSQMOATTPQGSNTALK--MNSPQSSPGMNPQP 482
Db 333 --PAVPGNQM-PPRPFGSGDSIMHPSMNOSSIADRGYQMRNPOMFOYSSP-----QP 384
OY 483 TSMSPRIRMRPGVAGSRIRPSPQSPGSLHSPVGCSSSTGNSISYINSSIALQALSE 542
Db 385 GSALSPRQ-----PS-----GGQIHTGMG-----SYOONSMSGSPGOG 418
OY 543 GHVSLGSSLASPLDKMGNLQNSPVNMPPLSKMGSLSKDCFG-LVGEPEEGTTGAOE 601
Db 419 QYGRQ-GGYPRQPN-----YNALPNANYPSGAMAGGINPMAGCGMGPPIRYG--- 468
OY 602 SSCHPGEQKET-----NDPNLPPAVSSEKADGOSRLHDSKQTKLLQLLTTSQDM 652
Db 469 -TLPPGRKSHASMGNRPGYPMANMPQVSGMCPPPCGMNRKQTQETAVAMHVANSION 527
OY 653 EPSPLASLSLSTNDSTSLPGSGSTHGTSLKEKHIIHLRLQDSSSPVDL----- 703
Db 528 RPPGYPMN-----NOGMMGTGPYGGGINS-----MAGMINPGPPYSKMGTMANNSA 576
OY 704 -----AKLT-----AEATGKDLG--OESSSTAQSEFTTIOQEPVSPK 739
Db 577 GMAASPENMGIGDYKLTPTATKMNKADGTPKTESKSSKSSSTTINEITIKIYELGGEPE 636
OY 740 KENALLRYLLDKD-----TKDI-GUPEITPKLERLDSKTD 774
Db 637 RKMVVDRIYLAFTEEKAMGNTLPAVGRKPLDLRLYVSVEKEIGLTVYNNKKMKRELATN 696
OY 775 PASNTKLIAMTEKE-----EMSFEPGDQPSSELDNEEIIIDDLQNSQLPOLFPDY 825
Db 697 LNVGTSSSAASLKKQYIQLYAECKIERGEDPPDI---FAAADSKSQ-----FKI 747
OY 826 RPPGAPAGSVDDQ-----AIINDLMQLFAENSP--VVP-----VGAOKT 861
Db 748 QPSPSAGSGMGPOTPOSTSSMAEGDLKPRTPASTPHSIOPLPLPKMSKNSNGVIGIDA 807
OY 862 ALRISQSTFN-----NPRPG-----QLGRLL-PNQMLPLDITLQSPTAGPFI--- 902
```

```
Db 808 FNDGSDSTFOKRNKSMTPNPGYQPSMNTSDMMGRMSYEPRKD-PYGSNRKAP-GSDPFMS 865
OY 903 ---PPIRNSPYSVIPQGMGNQGMIGNQ-----GMLGSSSTGMIGN 942
Db 866 GGGPNGMGDPYSRAAGRP-LGNVAMGRORHYPYGGPYDRVTEPGIIGEGNMSTG---- 920
OY 943 SASRPTMSGEAWPQSSAARVYTCATTSAMNRPVOGCMIRNPAASIPMPSSSQPCOROTL 1002
Db 921 -----ARQPNL-----MPSNDSGM-YSPSKYPPQOQOQ 948
OY 1003 OSOVNMGIPSELEMMNGGFOYSQQAAPPNOTAPWPESTILPIDQASFASQNRQPFSSSPD 1062
Db 949 OQO-----RHDSYGNQFSTQGTGTPSG-----SPPSQQTMYOQOQOQYKRPMDG 992
OY 1063 LLCPPAESPDEGALLDQYLALRNPFGLEIDRALGIPFLVSQSAVDEQESSQDS 1122
Db 993 TYGP-PARK--HEG-----EMT-----SVPTSGQGQ--PQO----- 1019
OY 1123 NIMLBOKAPVFPQOYASQAOQMSYSPMODPNFTMGORPSYATLRMQPRGLRPTGLV 1182
Db 1020 ---QQLPPAOPQ-P-ASQOQAAQ-P--SPQOQVYNOYGNAYPATATATERRRAGCP---- 1068
OY 1183 QNQ-ENQD---RLQDHLQAQONRQPLM-----NOISNVNV 1216
Db 1069 QNQPFQFGDRVSAAPGTMNAQNMPPQMMGGPIQASAEVAQGTMMQGRNDMTYNYANR 1128
OY 1217 NLTLPVPTQAPINQMLAQOREILNOHLRQRMHQOQOYQORTLMMKGGCLMMTBSM 1276
Db 1129 QST---GSAPQGP--AYHGVNRTDEML--HTQORANHEGS-----WPSGTROP-- 1171
OY 1277 VAPSGMPATMSNPRIPQANAOQFPPPNYGISQODDPGTGATTPQSLMSP---RMAAT 1333
Db 1172 YQPSAPVPRMTNP--PPSYQPPPSKQNH-ITQVSSP-----APLPRMENTTSFS 1219
OY 1334 QSPM---MQOSQANP---AYQAPSDI 1353
Db 1220 KSPFLSGKMKQKAGPPVASHIAPV 1247

RESULT 12
PASI_MOUSE STANDARD: PRT: 874 AA.
AC P97481; 008787; 055046;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Endothelial PAS domain protein 1 (EPAS-1) (HIF-1 alpha-like factor)
DE (MHLF) (HIF-related factor) (HRF).
GN EPAS1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain.
RX MEDLINE=97152468; PubMed=9000051;
RX "Tian H., Mcknight S.L., Russell D.W.;
RX "Endothelial PAS domain protein 1 (EPAS1), a transcription factor
RX selectively expressed in endothelial cells.";
RX Genes Dev. 11:72-82(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Hypothalamus, and Skeletal muscle;
RX MEDLINE=97272213; PubMed=9113979;
RX Ema M., Taya S., Yokotani N., Sogawa K., Matsuda Y.,
RX Fujii-Kuriyama Y.;
RX "A novel bHLH-PAS factor with close sequence similarity to hypoxia-
RX inducible factor 1alpha regulates the VEGF expression and is
RX potentially involved in lung and vascular development.";
RX Proc. Natl. Acad. Sci. U.S.A. 94:4273-4278(1997).
RN [3]
RP SEQUENCE FROM N.A.
```



```

Db 693 ----ARGPYMSPA-----MIALSNKTLKROQEYE-----E 720
Oy 820 QLPPTTRGAPAGSVKDAIINDLMQUTRANSPVTPVCAQKATLRISOSTFNNPRGQLG 879
Db 721 QAFODTSGDDPGTSSHLMMRMKSLMGCTCLPMP-----D 757
Oy 880 RLPLPQNLPLDTLTQSPGAGPFPPIRNSSPSVIPQPMGNOGIGNOG-----930
Db 756 KTISSANMAPDEFTQKSMKLG--QPLRLPLP-----PQPTSTSSGEMNATGFPPOCYASQ 811
Oy 931 --NLGNSVTGMIGNSAREPTMPSGEMAPQSAVRYVCAATTSAMNRPVYG 978
Db 812 FQDYPGPPGAQKVSGVASRLGSPFEPYLLPELTRYDC-----EVANPVAG 856

RESULT 13
PAS1_HUMAN STANDARD: PRT; 870 AA.
AC 099814; 099630;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Endothelial PAS domain protein 1 (EPAS-1) (Member of PAS protein 2)
DE (MOP2).
GN EPAS1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97152468; PubMed=9000051;
RA Tian H., McKnight S.L., Russell D.W.;
RT "Endothelial PAS domain protein 1 (EPAS1), a transcription factor
RT selectively expressed in endothelial cells.";
RL Genes Dev. 11:72-82(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Hepatoma;
RX MEDLINE=9079689;
RA Hogenesch J.B., Chan W.K., Jackiw V.H., Brown R.C., Gu Y.-Z.,
RA Pray-Grant M., Perdev G.H., Bradford C.A.;
RT "Characterization of a subset of the basic-helix-loop-helix-PAS
RT superfamily that interacts with components of the dioxin signaling
RT pathway";
RL J. Biol. Chem. 272:8581-8593(1997).
CC -1- FUNCTION: TRANSCRIPTION FACTOR INVOLVED IN THE INDUCTION OF OXYGEN
CC REGULATED GENES. SPECIFICALLY RECOGNIZES AN 8 BP HYPOXIA RESPONSE
CC ELEMENT (HRE). REGULATES THE VASCULAR ENDOTHELIAL GROWTH FACTOR
CC (VEGF) EXPRESSION AND SEEMS TO BE IMPLICATED IN THE DEVELOPMENT OF
CC BLOOD VESSELS AND THE TUBULAR SYSTEM OF LUNG. MAY ALSO PLAY A ROLE
CC IN THE FORMATION OF THE ENDOTHELIUM THAT GIVES RISE TO THE BLOOD
CC BRAIN BARRIER. POTENT ACTIVATOR OF THE TIE-2 TYROSINE KINASE
CC EXPRESSION.
CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER
CC BHLH PROTEIN, HETERODIMER WITH THE ARNT PROTEIN.
CC -1- SUBCELLULAR LOCATION: Nuclear (potential).
CC -1- TISSUE SPECIFICITY: EXPRESSED IN MOST TISSUES, WITH HIGHEST LEVELS
CC IN PLACENTA, LUNG AND HEART. SELECTIVELY EXPRESSED IN ENDOTHELIAL
CC CELLS.
CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS. HIGHEST TO HIF-1 ALPHA.
CC -1- SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC -----
DR EMBL: U01984; AAB41495.1; -
DR EMBL: U51626; AAC51212.1; -
DR TRANSFAC: T02718; -
DR MIM: 603349; -
DR InterPro: IPR003015; HLH_MYC.
DR InterPro: IPR001092; HLH_dlm.
DR InterPro: IPR001610; PAC.
DR InterPro: IPR000014; PAS.
DR Pfam: PF00785; PAC; 1.
DR SMART: SM00989; PAC; 2.
DR SMART: SM00353; HLH; 1.
DR SMART: SM00086; PAC; 1.
DR SMART: SM00091; PAS; 2.
DR PROSITE: PS00038; HELIX_LOOP_HELIX; 1.
DR PROSITE: PS50112; PAS; 2.
DR Repeat: DNA-binding; Nuclear protein; Transcription regulation;
DR Activator; Angiogenesis; Developmental protein.
FT DNA_BIND 15 27
FT DOMAIN 28 68 BASIC DOMAIN.
FT DOMAIN 84 154 PAS 1.
FT DOMAIN 230 300 PAS 2.
FT DOMAIN 304 347 PAC.
FT DOMAIN 474 480 POLY-SER.
FT CONFLICT 60 60 A -> E (IN REF. 1).
FT CONFLICT 539 539 G -> D (IN REF. 1).
FT CONFLICT 601 601 R -> H (IN REF. 1).
FT CONFLICT 693 693 N -> D (IN REF. 1).
FT CONFLICT 716 716 K -> E (IN REF. 1).
FT CONFLICT 722 722 P -> L (IN REF. 1).
FT CONFLICT 765 765 L -> F (IN REF. 1).
FT CONFLICT 769 769 S -> P (IN REF. 1).
FT CONFLICT 844 844 R -> C (IN REF. 1).
FT CONFLICT 847 847 K -> N (IN REF. 1).
SQ SEQUENCE 870 AA; 96425 MW; 3DF5B7B13AEC871D CRC64;

```

Query Match 3.6%; Score 278.5; DB 1; Length 870;  
 Best Local Similarity 20.1%; Pred. No. 8.9e-06;  
 Matches 207; Conservative 140; Mismatches 394; Indels 289; Gaps 44;

```

Oy 10 DSRARETRKRCPCPOLGSPKRNTEKRRNEDEKNYIEELA-ELIFANPDIONFNFKPD 68
Db 6 EKRRSSSEKRE-----KSRDAARCRKSRKTEVFYLAHELPLPH-----SVSSHD 52
Oy 69 KCAILKEVYKQIRQIKEDKAAANIDEVOKSDVSSTGGVYIDKALGFMLEALDGEFF 128
Db 53 KASIMRLAISFLRTKLLSSVCSSESEAEAD-----QOMDLVYKALEGFIA 100
Oy 129 VVNLGNAVVESENTQYLRVNOELMKNSVYSLHVGHTFEVYNLPLKSLVNGSGMSG 188
Db 101 VVTOGDMLFLESENISKPEGLTQVELTGHSIDFTPHCEIRNL---SLKNSGFG- 156
Oy 189 EPPRRNSHTFNCRMVLVPLPDSSEEGHDQEAHQKVEFTMQCAVAPQPSIKKEEG-----242
Db 157 -----KSKDMSTRDPPFMKACVTVTRGRVYNLKSATFWKL 193
Oy 243 -----EDLQSLICVARRVPKRERPLP-SSESFTTRDLOGK 279
Db 194 HCTGVKVVNCPNPNHSLGCKYKEPLSLCILIMCEPIQHSMDIPDLSKTFLSRSMQK 253
Oy 280 ITSLSDTSTRAAMKRGW--EDLVKRCIQFNAHQEGESVYAKRHNNHVLROGLAFSQY 337
Db 254 FYICDD--RTITELGYHHEELGSAVEFYALDSENNMT--KSHQNLCTKGQVVSQY 307
Oy 338 FRSLSGTLVAQOTKSKLRSQTNBPOLVSLH--MLHREONVCVMPNDLTGOTMGRP- 394
Db 308 RMLAHGCVLWLETGVTIYNPRNLQPCIMCVNVVLSIEKNDDVYFSMDQY-ESLFRPH 366
Oy 395 ---LNPISNSPAHQALCSGNGQDMTSLSSNINPINCPRK-----QMGMPDG-----439
Db 367 LMAANSIFDSS-----GKG-AVSEKSNFLFKLKEPEELAQAPLPDPAITSL 414

```

```

QY 440 RFGSGGGMNHSVSGMATTPOGSSNVALKMNPS--OSSPGMNPQOPTSMLSPRHMSPGVAG 498
D 415 DFGQNFEESSAVGKALLPSPQWATLRSHSTQSEAGSLPFTV-----PAAA 464
QY 499 SPRIPSPQSPAGSLHSPVCSSTGNSHSTYNSLNALQALSPGHVSGSSLASDGLK 558
D 465 PGSTTPATSSSS-----CSTPNSPEDYYSLLDNLK-----IEYIEKLFADTE 510
QY 559 MGNLQNSPVNMPPLSKMGSLSKDCFGLYGEPSEGTGQAESSCHPGQKETNDPNLP 618
D 511 AKDQCSGTQTFENLEDTLAPY-----IPMDGEFQLSPIC--PEERLLENPOST 559
QY 619 PAVSSEADQOSRLHDSKGT-----KLLQLLTSSQDMESPLASS 660
D 560 PQ-----HCFSAANTNIFQPLAPVAPHPSPFLDLFOOLESKTEPERRPMSSI 607
QY 661 LSPDNKSDTSLP---GSGSTHGTSLKEKHILHRLLDSSSPVDLAKLTAATGKLSQ 717
D 608 FPDAG--SKASLPCCGQASTPLSSMGGRN-----TQMPDPPLHFGPTKMAVG--DQRT 659
QY 718 ESSSTAGSEVTIKQEPVSPKRENALLRYLLDKDPTKDIGL--PEI--TFLERLDSKT 774
D 660 EFLGAAP-----LGPVSPHVSTFKTR-----SAKFGARGPNVLSAMVAL----- 702
QY 775 PASNTKILAMKTEKEEKSFEF---GDQPGSELNLE--FIIDDLQNSQLPOLFPDTPRGAR 830
D 703 --SNKLLKRLQLEKQAFDPGSGDPPGSGTSHLMKRRKNLRLGSGCP--LMPD----- 753
QY 831 AGSVKQAIINDLMQLTAENSPTYPVGAQKLTALRISOSTFNNRPQGLRLPQNLPLD 890
D 754 -----KPLSANPNNDKLTIONSMRG-----LGH--PLRLHLPLP 783
QY 891 ITTQSPGTAGP-----FPPIRNSPYSVIPQPGMNGMIGNOGNLNLSGTMGINS 943
D 784 ---QPSAISPGENSKSRFPQCVATQYQ-----DYLSSAHMKVSGM 822
QY 944 ASRPTMPSGE 953
D 823 ASRLGPSFE 832

RESULT 14
P300_HUMAN
ID P300_HUMAN STANDARD; PRT; 2414 AA.
AC Q09472;
DT 15-JUL-1998 (rel. 36, Created)
DT 15-JUL-1998 (rel. 36, last sequence update)
DT 01-MAR-2002 (rel. 41, last annotation update)
DE E1A-associated protein p300.
GN p300 OR P300.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95011587; PubMed=7523245;
RA Eckher R., Ewen M.E., Newsome D., Gerdes M., Decaprio J.A.,
RA Lawrence J.B., Livingston D.M.;
RT "Molecular cloning and functional analysis of the adenovirus E1A-
RT associated 300-kD protein (p300) reveals a protein with properties of
RT a transcriptional adaptor."
RL Genes Dev. 8:869-884(1994).
CC -!- FUNCTION: PROBABLE TRANSCRIPTIONAL ADAPTOR REQUIRED FOR THE
CC ACTIVITY OF CERTAIN COMPLEX TRANSCRIPTIONAL REGULATORY ELEMENTS.
CC MAY HAVE A FUNCTION IN CELL CYCLE REGULATION. BINDS TO AND MAY BE
CC INVOLVED IN THE TRANSFORMING CAPACITY OF THE ADENOVIRUS E1A
CC PROTEIN.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: CONTAINS 1 BROMODOMAIN.
CC -!- SIMILARITY: CONTAINS 1 ZF-TYPE ZINC FINGER.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U01877; AAA18639.1; -.
DR TRANSFAC: T01427; -.
DR MIM: 602700; -.
DR InterPro: IPR001487; Bromodomain.
DR InterPro: IPR003101; KIX.
DR InterPro: IPR000197; TAZ_finger.
DR InterPro: IPR000433; ZnF_ZZ.
DR Pfam: PF00439; bromodomain; 1.
DR Pfam: PF02172; KIX; 1.
DR Pfam: PF02135; zf-TAZ; 2.
DR Pfam: PF00569; ZF_1.
DR PRINTS: PR00503; BROMODOMAIN.
DR SMART: SM00297; BROMO; 1.
DR PROSITE: PS00633; BROMODOMAIN_1; 1.
DR PROSITE: PS50014; BROMODOMAIN_2; 1.
DR PROSITE: PS01357; ZF_ZZ_1; 1.
DR PROSITE: PS0135; ZF_ZZ_2; 1.
DR Transcription regulation; Nuclear protein; Bromodomain; Cell cycle;
DR Zinc-finger.
KW
FT DOMAIN 11 17 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 1067 1139 BROMODOMAIN.
FT DOMAIN 1572 1818 BINDING REGION FOR E1A ADENOVIRUS.
FT ZN_FING 1664 1707 ZF-TYPE.
FT DOMAIN 797 800 POLY-SER.
FT DOMAIN 1519 1526 POLY-GLN.
FT DOMAIN 2066 2069 POLY-GLN.
FT DOMAIN 2190 2195 POLY-GLN.
SQ SEQUENCE 2414 AA; 264143 MW; 6BFF909EE4B9D693 CRC64;

Query Match 3.6%; Score 277; DB 1; Length 2414;
Best Local Similarity 18.9%; Pred. No. 3.8e-05;
Matches 222; Conservative 117; Mismatches 363; Indels 474; Gaps 50;

QY 381 VNPNDLTGCTMGKPLNDI-----SSNSPAHQALCSGNPGDMLTSSINPEPING 429
D 47 LINSTELGLNGLGDINDQLQTSIGMVODAAASKHRLSELLRSQS-----SPILNMGVGG 99
QY 430 PKQMGMPMRFGSGGMNHSVSGMATTPOGSSNVALKMNPSPOSSPGM-----NPGQ 481
D 100 PGQVMA-----SOAQOSSPQLGLINSVKSPPMT 127
QY 482 PTSLSPRRHRMSPGAVSPRIPSPQSPAGSLHSPV---GVCSSTGNSHSTYNSLIN--AL 537
D 128 QAGLTSP--NMGGTSG--PQGGPTQ--STGMNNSPVNQPAMGNNTG-----TAAGNPNCA 177
QY 538 QALSEGHV-----SLGSSLASPDLK-----MGNLQNSPVNMPPLS----- 575
D 178 IAAAGNCGGIMPVNMVNGSIGAGRGDMQYPNMGKSAGNLLEPLQOGSPQMGQGTGLR 237
QY 576 -----KMGSKDKDCGIVGEPSEGTGQAESSCHPGQKETN---DPNLPVAVSSERA- 626
D 238 GPPLPKMGKMNPN---PYGSPYTONPGQIGASGLQLOTIVLVSNNLSPPAMDKKAV 294
QY 627 --DGOSRLHDSKGTQLQLLTTRKSDMEP---SPLASSLSDTNKSDTSLSPGSGTGTGT 681
D 295 PGGGMNMM---GQAPAPV-----QDPGLVTPVAQ-----GMGGAHT 329
QY 682 SLKEKHILHRLI-----QDSSSPVD-----LAKLTAATGKLD-- 714
D 330 ADEKRLIQOQLVLLHAHKCORREANGEVNOCMLPHCRTKKNVNLNMTHTQSGSKSQ 389
QY 715 -----LSGSSSTAGSEVTIKQEPVSPKRENALLRL---LQDKDTKDIGLPE 761
D 390 VAHCASSROIISHMKNCTRHDCPVCLPLKNAGDKRNOQDPLTGAPVGLGNPSSLAGV--QO 448

```



```

QY 762 ITPKLELRSKTDPAANTKLA-----MKTE-----KEEMSPGQDQGS--ELD 804
DB 449 SAPNLSVY-SQIDPSIRAYALALGPQVYNMPTQPOVQAKNOQPGSGPQGMREKS 507
QY 805 NIEEHLDLQNS--QLPQLEPPTPRGAPAGSVQKAIINDLMQJTAENSPVTPVGAOKT 861
DB 508 NMSASPMGVNGVGQVTPSLSDSM-----LHSAINSQNPMMSENAVSPLGPMPT 558
QY 862 AL-----RISQSTFNNPRPGQLG-----879
DB 559 AAQPSITGIRKQWHDITQDLRNHLVHLYQAIPTPPAALKDRRMENLYAVARKVGD 618
QY 880 -----RLTPNQNLPLDITLOS-----PTGAGPPPIRRSSPYVIP 915
DB 619 MYESANNRAEYTHLAEKTYIKQKELEKRRTRLOKQNMPLPAAGMNVYSNMPGNMGOP 678
QY 916 QPMKMGNOGMIGNOGNLGNSSTGMTIGNSASRPTMPSGEMAPQSSAVRYTCAATTSAMNRP 975
DB 679 QPMGMSN-----GLPDPDS--MIRGSVPNQMP-----704
QY 976 VQGGIRNPASIPMRSSQGOQROTLOSQVYNITPSELEMMNGGPQYSQQAAPPNOTAP 1035
DB 705 -----RITPQSGLNQFG--QMSMAQPPVPRQTPPLQ---734
QY 1036 WPESILPIDQASFASQNRQPGSSPDDLCPHPAESPSDEGALLDQYLALRNFDGLEE 1095
DB 735 -----HHQQLAQPGALNRP--MGYGRMQQPSNQGQFLPQ-----767
QY 1096 IDRALGIPELVYSQASQVDEQFSSQSDSNIMLEQKAPVEPQYQASQAOAAGS--YSPMQ 1152
DB 768 -----TOFPSSQGMVNTNPLAPSSGQAPVSOQAOSSSCPVNSPIM 808
QY 1153 DPNHTMGQBSYATLRMQRPGLRPTGLVQNPQNLQQLQHLQQAQONQPLMNQSN 1212
DB 809 PP-----GSQGSHTHCPQLPQPALH-----QNSPSP-----834
QY 1213 VSNVNLTRPGVPTQAPINAOMLAQROREILNQHLRQRMHQOQVOQRTLMNRQGLNM 1272
DB 835 -----VPSKTP-----TPH-----HTPSSI-----GAQO 853
QY 1273 TPSSVADSGMPATMSNPRIQANA-----QQFPPPNYGISQQDPDGTGATTPOSPLMS 1327
DB 854 PRATTIPARVPTPRAMPGPQSOALHPPRPQTPPTQLPQOVQPSLPAAPSADQPOQ 913
QY 1328 PRMAHT-----QSPWMOQSQAANPAPQAPSDING 1355
DB 914 PRSQOSTRAASVTPPVAPLLPPOPATPILSQPAVSIEG 949

RESULT 15
SIML_HUMAN
ID SIML_HUMAN STANDARD; PRT; 766 AA.
AC P81133;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Single-minded homolog 1.
GN SIML.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=97343329; PubMed=9199934;
RA Chast R., Scott H.S., Chen H., Kudoh J., Rossier C.,
RA Mioshima S., Wang Y., Shimizu N., Antonarakis S.E.;
RT "Cloning of two human homologs of the Drosophila single-minded gene
RT chromosomal region."
RL Genome Res. 7:615-624(1997).
-i- FUNCTION: TRANSCRIPTIONAL FACTOR THAT MAY HAVE PLEIOTROPIC EFFECTS

```

```

CC CC DURING EMBRYOGENESIS AND IN THE ADULT.
CC -i- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER
CC BHLH PROTEIN. HETERODIMER OF SIML AND ARNT.
CC -i- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS.
CC -i- SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.
CC -i- SIMILARITY: CONTAINS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC CC
CC DR EMBL; U70212; AAB62395.1; -.
CC DR MIM; 603128; -.
CC DR InterPro; IPR003015; HLH_MYC.
CC DR InterPro; IPR001092; HLH_dim.
CC DR InterPro; IPR001610; PAC.
CC DR InterPro; IPR000014; PAS.
CC DR Pfam; PF00785; PAC; 1.
CC DR Pfam; PF00989; PAS; 2.
CC DR SMART; SM00353; HLH; 1.
CC DR SMART; SM00086; PAC; 1.
CC DR SMART; SM00091; PAS; 2.
CC DR PROSITE; PS00036; HELIX_LOOP_HELIX; 1.
CC DR PROSITE; PS0112; PAS; 2.
CC KW Developmental protein; Neurogenesis; Nuclear protein; Repeat;
CC FT Transcription regulation; DNA-binding.
CC FT DNA_BIND 1 13 BASIC DOMAIN.
CC FT DOMAIN 14 54 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
CC FT DOMAIN 77 147 PAS 1.
CC FT FT 218 288 PAS 2.
CC SQ SEQUENCE 766 AA; 85474 MW; 6D33694BF9A29DD6 CRC64;

Query Match 3.6%; Score 273.5; DB 1; Length 766;
Best Local Similarity 20.9%; Pred. No. 1.3e-05;
Matches 168; Conservative 112; Mismatches 292; Indels 233; Gaps 36;

QY 32 NTERKRENEQNKYLEELIFANFNIDNPNFPDCAIKETVKQIR-QIKQEKAA 90
DB 6 KNAATREKENSEFEYELAKLPL-----ASAITSYVDKASIIIRLTSTLKARVVPFEG 61
QY 91 AANIDEVQKSDVSTGQGVIDKDALGPMLEALDGFVNVNLEGNVVSENVTOYLRYN 150
DB 62 EAWGHSSRTSPLDNVR-----ELGSHLQTLQDGFIVVAPDKIMVISTEASVHLGLS 115
QY 151 QBELMKSIVYSLIHVGHTEFEVKNILPKSIYNGSGWSGEPFRNSHTPCMLYKPLPDS 210
DB 116 QVELTGNISIVEYIHPADHDEMTAVL-----140
QY 211 EEEGHQNEAH--OKYE-----TMOCFAVSOPRSIKEBEDLQSC-----248
DB 141 --TAHQPYHSHFVQYETELERSFEFLRMKCYLAKRNAGLCGGKYVHCGYIKIRQYSLDM 198
QY 249 -----LICVARRVPKKERFVLPSS-----SESEFTTRDLOGKITSLDTSTM 289
DB 199 SPFGQCYQNVGLVAVGHSLP-----PSAVTEIKLHNMFMFASLDMKLIFLDS---R 248
QY 290 AAMKPGWE--DLVRRCLQKFLPAQHEGESVYAKRHHNEVLRQGLAFISOIYRFSLSDTGLV 347
DB 249 VAEITGYEPDILLIEKTL--YHNVH-GCDTFLRCANHLLLVKGVQVTTYVYRFLKHHGGMV 305
QY 348 AAQTKSKILRSQTNPEQPLVSLIMHLHREQVNCVANNPDLTQGTMKPLNPITSSNPAQDA 407
DB 306 WVQSTATTIVHNSRSSRPRHCIVS-----NYVLTDYEYKILQLS--LDQISASKPAFVS 356
QY 408 LCSGNPGQDMTLSSNINPPIGPKQMGMPGRFGSGGMHVGGMQATT--POGSNV-A 464
DB 357 TSSSTP-----TMTDN-----RKGAKSRLSSSKSKSTSPYQYSGHT 395

```



```

Db 606 TKA-----RTQDLALAPKQ00000VGTNTQ000LAVATA00000QLTAAALQR 658
QY 947 PTMP-----SGEMAPQSSAVRVTCAATTSAMNRPVQGGMRNPASISIPMRPSSQGGQRTL 1002
Db 659 PGAPVNMHNGVQVPAASV-----STQTAQNSILKAKMRKQO--FVRA-----LATL 706
QY 1003 QSOVMMI--GPSELEMMNGDPQYSSQ-----QAPNQTAPWPESTLPIQDASFASQNR 1053
Db 707 KTEIGVAGQNKVYGHLLTQQ0000ATNLQOVNAAGNKVMVMTGTGPI-----TLQNG 761
QY 1054 QPSSSDLLCPHRAESPSDECALLDQLYALRNFDGLEDLRAIGIPELVSSQSAVD 1113
Db 762 QTL-----HAATAAGVDCQ000LQLFQKQ011000QMLQ001A10M00Q0QAAYQ 811
QY 1114 PEQFS-----SODSNIMLEOKAPVFPQYA--SQAMAGSVSPMDPNFHTMGORSPATYL 1168
Db 812 AQQ00000VQ0000VNA0000A0000A0000A00000-----REQ000VQAQQA 860
QY 1169 RMPQRPGLRPTGLVQNPQDLRLQLHRLQAQNRQPLMNOISNVSVNLTLPBGVPTQA 1228
Db 861 QHQ0ALANATQQLQVAPNPFITSHQ00000QLHNLQ00LQ0-----QAQA 908
QY 1229 PINAQMLAQORELNLHQRQMHQ00Q0VQ0RTLNR-----GQ-GL 1270
Db 909 QVQAQVQAQQAQ00-----Q0QREQ00QNIQ0LYVQ0SGATSQ00Q0HQSGQLQL 961
QY 1271 NMTPSVAPSGMPATMSNPRIPOANAOFPFRPVYGISQ0PDPGFGTATTPQS----- 1323
Db 962 SSVPSFVSSTTTPA-----GIATSSALQALASASGALFQAKPCTGSSSTSTSVVITTN 1016
QY 1324 ----PLM--SPRMAHTQSPMMAQSQANPAPYQAPSDINGMAQNMGNMFSQ0SPPHFCQO 1378
Db 1017 QSSPPLTSSSTVASIQQAQ0QSAQVHQHQLIS-----ATINGGQQQ0Q0GSPSLPPT 1069
QY 1379 ANTSMSNMANNINVSATNTGMSNMAMQGISMTSVTS-----VSTGSLSMG 1428
Db 1070 TNPILAMTSM-----MNAATGHLSTAPPVYVSTAVTSSPQLVLLSTASSGCGGSGIP 1124
QY 1429 PEQVNDPALRG--GNLEPNQLPGMDMKQESDPTRK 1462
Db 1125 ATPPKETPSKGPATVLPICSPKTPVSGKDCITTPK 1160

```

```

RESULT 17
HIFA_HUMAN
ID HIFA_HUMAN STANDARD; PRT; 826 AA.
AC Q16655;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypoxia-inducible factor 1 alpha (HIF-1 alpha) (ARNT interacting protein) (Member of PAS protein 1) (MOP1) (HIF1 alpha).
GN HIF1A.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 166-170; 259-289 AND 771-781.
RX MEDLINE=95296340; PubMed=7539918;
RA Wang G.L., Jiang B.-H., Rue E.A., Semenza G.L.;
RT "Hypoxia-inducible factor 1 is a basic-helix-loop-helix-PAS heterodimer regulated by cellular O2 tension."
RL Proc. Natl. Acad. Sci. U.S.A. 92:5510-5514(1995).
RP [2]
RN SEQUENCE FROM N.A.
RC TISSUE=Hepatoma;
RX MEDLINE=97236817; PubMed=9079689;
RA Hogenesch J.B., Chan W.K., Jackiw V.H., Brown R.C., Gu Y.-Z.,
RT Pray-Gant M., Perdev G.H., Bradfield C.A.;
RT "Characterization of a subset of the basic-helix-loop-helix-PAS superfamily that interacts with components of the dioxin signaling pathway."

```

```

RL J. Biol. Chem. 272:8581-8593(1997).
RN [3]
RP SEQUENCE FROM N.A.
RA Rupert J.L., Hochachka P.W.;
RT "HIF1a sequence in the quechua, a high altitude population."
CC Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: INVOLVED IN THE INDUCTION OF OXYGEN REGULATED GENES. SPECIFICALLY RECOGNIZES AN 8 BP HYPOXIA RESPONSE ELEMENT (HRE).
CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER BHLH PROTEIN. HETERODIMER OF AN ALPHA AND A BETA (ARNT) SUBUNITS. THE ALPHA SUBUNIT IS UNIQUE TO HIF-1 WHEREAS HIF-1 BETA (ARNT) CAN DIMERIZE WITH OTHER BHLH-PAS PROTEINS. INTERACTS WITH HSP90.
CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -1- TISSUE SPECIFICITY: EXPRESSED IN MOST TISSUES. HIGHEST LEVELS IN KIDNEY AND HEART.
CC -1- INDUCTION: UNDER REDUCED OXYGEN TENSION.
CC -1- DOMAIN: BOTH HYPOXIA RESPONSIVENESS AND TRANSACTIVATION CAPABILITY RESIDE WITHIN THE C-TERMINAL PART.
CC -1- PTM: REQUIRES PHOSPHORYLATION FOR DNA-BINDING.
CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF TRANSCRIPTION FACTORS.
CC -1- SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC
DR EMBL: U23431; AAC50152.1; -.
DR EMBL: U29165; AAC51210.1; -.
DR EMBL: AF207601; AAF20139.1; -.
DR EMBL: AF207602; AAF20140.1; -.
DR EMBL: AF208487; AAF20149.1; -.
DR TRANSFAC: T01610; -.
DR MIM: 603348; -.
DR InterPro: IPR003015; HLH_Myc.
DR InterPro: IPR001092; HLH_dim.
DR InterPro: IPR001321; Hypoxindf1A.
DR InterPro: IPR001610; PAC.
DR InterPro: IPR000014; PAS.
DR Pfam: PF00785; PAC; 1.
DR Pfam: PF00989; PAS; 2.
DR PRINTS: PR01080; HYPOXIA1F1A.
DR SMART: SM00353; HLH; 1.
DR SMART: SM00086; PAC; 1.
DR SMART: SM00091; PAS; 2.
DR PROSITE: PS00038; HELIX_LOOP_HELIX; 1.
DR PROSITE: PS50112; PAS; 2.
KW Repeat; DNA-binding; Nuclear protein; Transcription regulation;
KW Activator; Phosphorylation.
FT DNA_BIND 17 30 BASIC DOMAIN.
FT DOMAIN 31 71 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
FT DOMAIN 85 158 PAS 1.
FT DOMAIN 228 298 PAS 2.
FT DOMAIN 302 345 PAC.
FT DOMAIN 615 621 POLY-THR.
SQ SEQUENCE 826 AA; 92670 MW; ABDAF7DMA135BE2D CRC64;

```

Query Match 3.58; Score 269.5; DB 1; Length 826;

Best Local Similarity 19.38; Pred. No. 2.1e-05;

Matches 186; Conservative 150; Mismatches 337; Indels 289; Gaps 41;

```

QY 1 MSGGENTSPSRAETKRREKCPQGLSPKRNTERKRNREQENYIELELIFANRNDI 60
Db 1 MEGAG-CANOKKTKISSRKE-----KSRDAARSRRKSEVYELAHQPLPHNVS 51
QY 61 DNFFKPKDCAILKEVYKQIROIKEQEKAAANIDEVQKSDVSTGCGVIDKALGPMML 120

```



FT DOMAIN 1794 1799 POLY-GLN.  
 FT DOMAIN 1856 1863 POLY-GLN.  
 FT DOMAIN 2044 2059 POLY-PRO.  
 FT DOMAIN 2405 2408 POLY-ALA.  
 FT DOMAIN 3216 3220 POLY-PRO.  
 FT DOMAIN 3380 3409 POLY-GLN.  
 FT DOMAIN 3412 3420 POLY-GLN.  
 FT DOMAIN 3534 3550 POLY-GLY.  
 FT DOMAIN 3620 3623 POLY-PRO.  
 FT DOMAIN 3659 3662 POLY-SER.  
 SQ SEQUENCE 3726 AA: 406567 MW: 913ACBE58A72C98 CRC64:

Query Match 3.5%; Score 268.5; DB 1; Length 3726;  
 Best Local Similarity 19.5%; Pred. No. 0.00016;  
 Matches 285; Conservative 170; Mismatches 574; Indels 433; Gaps 65;

Oy 32 RNTERRNREQENKYLE-ELAEILIFANFDIDNFRPKDCAILKETVQIRQIKEDQKAA 90  
 Db 2301 ROKARKNENQEGKDEGERELTNDRIYRTSNLYQCKKSLVFQRIFDL--IKHQKIC 2358  
 Oy 91 AANDVQKSDVSTGCGVIDKAL-----GPMLEALDGFVVNLEGNVAV 139  
 Db 2359 YKDEDEEGODD---SQNEDSMAMEILPTSSSCSTPMPSQAYS-----TPAPSA 2408  
 Oy 140 SENVTQYLRNNOELMNKSVSYILVHGHTFEVKNLPRKSYVNGSGWSGEPRRNSHTEN 199  
 Db 2409 NTAPSAFQLTRETDELATFNKAKASDEK-----FR-----QADPPS----- 2446  
 Oy 200 CRMVLKPLPDESEEGHNDQEAHQKYEYTM-QCPAVSOPKSIKEGEDLOSCLICVARVPM 258  
 Db 2447 ---AOPNQTQKOGOPKPEMOQOLEQKTNAPQPKLPAPASL-----PQPPQ 2495  
 Oy 259 KERPLPSESESTTQODLG-KITSLSDTST-----MRAMPGMEDLYRCIOKFAH 311  
 Db 2496 APPPCPLPQSSPSQSLHLPLKPLHTSTPQOLANLPPOLLPYOCDOCKLAFPEEHQ 2555  
 Oy 312 EGESYVAKRHHHEVLROGIAFSQYRFLSD--GTLVAOTKSKLIRSQTNNEPOLVTS 369  
 Db 2556 EHQQLHFLSAQNOFIHPQFLDLSLMPFLPSPNPLASQLSGAI-----PQIPAS 2608  
 Oy 370 LHMILREONVCVNDLTGTQTMGKPLNPISNSPAHQALCSG--NPGQDMLTSSNINPFI 427  
 Db 2609 ---SATSPSTPTSTMTNLTKRLEKASASPGENDSGTGEPEQRDKRLRTTIT 2658  
 Oy 428 NCPKROMGMPKRGSGSG-----MNHVS---GMQATTQGSNTYALKMSPQSSGSG--- 476  
 Db 2659 ---PQOLEILYOKYLLDSNPTRKMLDHAHEVGLKRRVQ---VWFQNTRARERKQGR 2711  
 Oy 477 -MNPQPTSMLSPRHRMSP-----GVAGSPRIPPSQFSPA----- 510  
 Db 2712 AVGPQO-----AHRRCRPFCALEKATKATLAHIRSRMHAKRAGVYLTLSAMLLDCD 2764  
 Oy 511 GSLHSPVGCSTGNSHSTYNSLNAALALSGHGVSIG-----SSLASP-DLKM 559  
 Db 2765 GGIQKKGDIIFDQTSFSSHLPSS-----SDQGVPLSPVSKTMELSPTLLSPSSIKV 2816  
 Oy 560 GNLOK-----SPVMNMPPLSKMGLSDKDFGLGEPSEGT---GQAESCHPGEK 610  
 Db 2817 EGIEDFESPMSSVNLN---FDQTKLNDCCSVNTAITDTTDTDEGNADDSATGAT 2872  
 Oy 611 ETND--PN--LPPAVSERADGOSRL-----HDSKQTKLLLTTSKQ 651  
 Db 2873 ETKSAPMEGLTKAMAMMSEVEDRLSGVSPASFSKEDNECTVYSE--TSSLADP 2931  
 Oy 652 MEPSPLASLSDTNKSDTSGSLPGSGSTGTGTSLEKHKHILHRLLODSSSPVDLAKLTAET 711  
 Db 2932 CSPSPGASGAGKSGDG--GDRPGOKRFRQMTNLTQLKYLKSCFNDRFP---TMECEVL 2987  
 Oy 712 GMDLSQESSSTFAPGESEVITIKQEPVSPKKKENALLKRLDKDTKDIGLEPIT---PKLE- 767  
 Db 2988 GNDIG-----LPRKRVQVWFQNNARAKEKSKL-----SMAKHFGINQTSYEGKTEC 3034

Oy 768 -----RLDSKTDPAENFKLIANKTEKEKSEPEGDPGSELNLEIIDLQNSQL 818  
 Db 3035 TLGGIKYSARLSVRHIFESQHIKVK-----DTISQDKKEKEYPDPAVRQL 3083  
 Oy 819 POLFPDTRGACAGSVCKAIIINDLMQLTAENSPLYTPVGAQXTALRISOSTNNRPPGOL 878  
 Db 3084 -----MAQOELDRIKKANENVGLAAO-----QQGMEDN----- 3111  
 Oy 879 GRLLPNQNLPLITITQSPGACGPPPIRNSPSYVIRPGGMGNQGMIGNQNLGSSNG 938  
 Db 3112 ---APLQALNLTPT-----YPLQGIIP---VLLPGL----- 3137  
 Oy 939 MIGNSARPTMPSGEMAPSSAVRVYCAATTSAMNRPVQGMIRN-----PAASIPAR 991  
 Db 3138 -----NRSLSLG-----FTPATLTALSKPRNLMGLPSTTVPSPLGTLSPRK 3180  
 Oy 992 PSSQPGQROTLOSQVANNIGPSELEMMNGPQYSSQOAPPNQTPAWDEST--LPIDQASFA 1049  
 Db 3181 PSSASLSPPTPAQATWAMAPQ-----PQOPQOPQPPVQOPPPPPAQQIPAPQITPQ 3233  
 Oy 1050 SQNRQPFSSPPDLCPHFAESPSDEGALLDQVIALNPFGLIEDIALGIPE----- 1104  
 Db 3234 QQRKDRGKGEKKEKRAHKGEP-----LVPKKEGE 3267  
 Oy 1105 -----LVSSQAVDPEQFS-----SODSNIMLEOK-----APVFPQOYASQA 1141  
 Db 3268 APPAGGTISAPLPMMEYAVNDPAQALQALTSPTALTTSQFLPYFPGSPYAPRI 3327  
 Oy 1142 QMA-QGSSYSPMODPNEFTMGORPSYATLRMPDPGLRPTGLVONQNLRLQJHRLQAO 1200  
 Db 3328 PALQSGY---LQPMWGMGLFP--YSPALSRPLMGISPSGLIQ-QYQYQOSLQEAIOQ 3382  
 Oy 1201 QNRQPLMANISVSNVNLTLRGCVPTQAPINQMAQORRELLNQLRORQHQQOQYQ 1260  
 Db 3383 QQQQQQQQQQ-----QQQQQQQ-----LQQQQQQQQQQQ 3413  
 Oy 1261 RFLMRGQGLNMTP---SWAPSGMPATWS-----NPRIPQANAOQFPPPYGISOQPD 1312  
 Db 3414 QQQQQQQQPKASQTPVPOGAASPDKPAKESPRPEQKNVRELSPLLPRP-----EERE 3468  
 Oy 1313 PGFTGATTPQ--SPLMSPRMAH 1332  
 Db 3469 AESKSASADSLCDPFTVPRVQY 3490

RESULT 19  
 BMAL\_HUMAN  
 ID BMAL\_HUMAN STANDARD; PRT; 583 AA.  
 AC 000327; 000313; 000314; 000315; 000316; 000317; 099631; 099649;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE BMAL1 protein (Brain and muscle ARNT-like 1) (Member of PAS protein 3)  
 GN BMAL1 OR ARNTL.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxId=9606;  
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING (ISOFORMS BMAL1 A-F).  
 RC TISSUE=Brain;  
 RX MEDLINE=97289529; PubMed=9144434;  
 RA Ikeda M., Nomura M.;  
 RT "cDNA cloning and tissue-specific expression of a novel basic  
 RT helix-loop-helix/PAS protein (BMAL1) and identification of  
 RT alternatively spliced variants with alternative translation  
 RT initiation site usage".  
 RL Biochem. Biophys. Res. Commun. 233:258-264(1997).  
 RP SEQUENCE FROM N.A. (ISOFORM MOP3).  
 RC TISSUE=fetal brain;  
 RX MEDLINE=97236817; PubMed=9079689;

RA Hogenesch J.B., Chan W.K., Jackiw V.H., Brown R.C., Gu Y.-Z.,  
 RA Pray-Grant M., Perlew G.H., Bradford C.A.;  
 RT "Characterization of a subset of the basic-helix-loop-helix-PAS  
 RT superfamily that interacts with components of the dioxin signaling  
 RT pathway";  
 RL J. Biol. Chem. 272:8581-8593(1997).  
 RN [3]  
 RP SEQUENCE FROM N.A. (ISOFORM BMAL1B).  
 RA Tian H., Russell D.W., McKnight S.L.;  
 RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP INTERACTION WITH CLOCK.  
 RX MEDLINE-98279137; PubMed-9616112;  
 RA Gekakis N., Staknis D., Nguyen H.B., Davis F.C., Wilsbacher L.D.,  
 RA King D.P., Takahashi J.S., Weitz C.J.;  
 RT "Role of the clock protein in the mammalian circadian mechanism";  
 RL Science 280:1564-1569(1998).  
 CC -1- FUNCTION: CLOCK-BMAL1 HETERODIMERS BIND TO AN E-BOX ELEMENT (3'-  
 CC CACGG-5') THEREBY ACTIVATING TRANSCRIPTION OF PER1, AND POSSIBLY  
 CC OF OTHER CIRCADIAN CLOCK PROTEINS.  
 CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER  
 CC BHLH PROTEIN. FORMS AN HETERODIMER WITH CLOCK. INTERACTS WITH  
 CC HSP90; WITH AHR IN VITRO, BUT NOT IN VIVO.  
 CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).  
 CC -1- ALTERNATIVE PRODUCTS: AT LEAST 7 ISOFORMS; BMAL1A (SHOWN HERE),  
 CC BMAL1B/AP3, BMAL1C, BMAL1D, BMAL1E, BMAL1F AND MOP3; ARE PRODUCED  
 CC BY ALTERNATIVE SPLICING.  
 CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN THE ADULT BRAIN, SKELETAL  
 CC MUSCLE AND HEART.  
 CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF  
 CC TRANSCRIPTION FACTORS. HOMOLOG OF DROSOPHILA CYCLE PROTEIN.  
 CC -1- SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; D89722; BAA19968.1; -;  
 DR EMBL; AB000812; BAA19935.1; -;  
 DR EMBL; AB000813; BAA19936.1; -;  
 DR EMBL; AB000814; BAA19937.1; -;  
 DR EMBL; AB000815; BAA19938.1; -;  
 DR EMBL; AB000816; BAA19939.1; -;  
 DR EMBL; U51627; AAC51213.1; -;  
 DR EMBL; U60415; AAB37248.1; -;  
 DR MIM; 602550; -;  
 DR InterPro: IPR003015; HLH\_Myc.  
 DR InterPro: IPR001092; HLH\_dim.  
 DR InterPro: IPR001067; NucleosLocator.  
 DR InterPro: IPR001610; PAC.  
 DR InterPro: IPR000014; PAS.  
 DR Pfam: PF00010; HLH; 1.  
 DR Pfam: PF00785; PAC; 1.  
 DR Pfam: PF00989; PAS; 2.  
 DR PRINTS: PRO0785; NCTNSLOCATR.  
 DR SMART; SM00353; HLH; 1.  
 DR SMART; SM00086; PAC; 1.  
 DR SMART; SM00091; PAS; 2.  
 DR PROSITE; PS00038; HELIX\_LOOP\_HELIX; 1.  
 DR PROSITE; PS00112; PAS; 2.  
 KM Repeat: DNA-binding; Nuclear protein; Transcription regulation;  
 KW Alternative splicing.  
 FT DNA\_BIND 30 42 BASIC DOMAIN.  
 FT DOMAIN 43 83 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).  
 FT DOMAIN 100 172 PAS 1.  
 FT DOMAIN 283 353 PAS 2.  
 FT DOMAIN 358 401 PAC.  
 FT VARSPLIC 1 4 MINI -> MADQRMDISITSDPMSPPPTDLLSSSLGTSGV

FT VARSPLIC 1 16  
 FT VARSPLIC 181 181  
 FT VARSPLIC 182 583  
 FT VARSPLIC 231 348  
 FT VARSPLIC 235 258  
 FT VARSPLIC 259 583  
 FT VARSPLIC 400 483  
 FT VARSPLIC 484 583  
 FT CONFLICT 26 26  
 FT CONFLICT 80 80  
 FT CONFLICT 130 130  
 FT CONFLICT 216 216  
 FT CONFLICT 221 221  
 FT CONFLICT 375 375  
 FT CONFLICT 470 471  
 SQ SEQUENCE 583 AA; 64206 MW; 2AAB7EBBA471119 CRC64;  
 Query Match 3.5%; Score 267; DB 1; Length 583;  
 Best Local Similarity 23.3%; Pred. No. 1,8e-05;  
 Matches 146; Conservative 89; Mismatches 245; Indels 146; Gaps 29;  
 QY 1 MSGNGENTSDP-----SRAETRRKECPDOLGSPKRNTERKREQENKYEELAEI 52  
 DB 4 IESMDTKDPDPRHGLEYTEHQGRIKNAREAHQS-----EKRRDKMNSFIDELASL 55  
 QY 53 IFANFNIDNFNFPDCKALIKETVQIRQIRK-EQKAAANIDENVQSVSTGCGVID 111  
 DB 56 V-----PTCNMMSRKLDLTVLRNAVQHMKTGRATNPTYEANKKPFPLSD----- 101  
 QY 112 KDALGPMLEBALDGFEEVNVILE-GNVFVSENVQVLYRNOEELMKNKSVSILHVGDATE 170  
 DB 102 -DELKHLILAAAGFLFVVGCDRKLILFESVSKILNYSQNDLIGSLDFDIYLPKDIAX 160  
 QY 171 FVKNLTPKS-----IVNGSWSGEPPRRNSHTNCRMLY-KPLPD 209  
 DB 161 -VKEQLSSSDTAPRERLIDAKTGLPVKTDITDTPGSRSLCGARR---SFCRMKCNPSVK 216  
 QY 210 SEEGCHDNOEAHQYEMQFAVSQPKSIK-----DEGDIQSCLICVA 253  
 DB 217 VEDKDPSTCSKKKADKKSCTIHSYGILKSWPPTKMGLEDDEPDNEGCL-SCLVAIG 275  
 QY 254 RR-----VPMK-ERPVLPSSESEFTTRDLOKITSLD-TSTRAMAKRGWEDLYRRCIOKF 307  
 DB 276 RLMSHVVPQPVNGEIRKMSMEYVSRHAIDGKPFVVDRAITAILAYLP-QELLGTSCEYEF 334  
 QY 308 HAQHEGESVYAKRHHEVL--RQGLAFSQTIRFSLSDGLVAQYKSKILRSQTTNEPQ 365  
 DB 335 HQ-----DDICHLAECHROVLOTREKIT-TNCFYFKIDGSEFTILRSWFSEFMMPWTEVE 389  
 QY 366 LVISLMLHLREQVYCVANNPDLTGOTMCKPLNPISNSSPAHALCSGPGQDMTLSSINF 425  
 DB 390 YIVS-----TNTVVLANVLEG--GDPYFPOLTAAP-----HSMDSMLPSGEG 430  
 QY 426 PINGPKEQMGMPAGREGSGGMNHSV--MQATTPGGSNYALKMNPSOSSPGMNPQOP 482  
 DB 431 PKRHTPVPPGILGTGRAGAKIGMAELMEIHLIRIS-----SPSSGSSPLNI 481  
 QY 483 TSMLSPRHRMSPG-----VAGSPRIPPQSPSPAGSLSPYGVCSSTGNSHSYNSLINAL 537  
 DB 482 TS--TPPDASSPGKILNGTDPDIPSSGLSGAQENP-----GYPSDDS--- 527  
 QY 538 QALSEGHGVSIL-----GSSLASPD 556  
 DB 528 SILGENPHGIDMINDOGSSSPSND 553



```

QY 801 -SELD-NLEETLDDLNSQLPOLFPDTRPGAPAGSVKQAIIT-----NDLM 844
DB 1507 DSEVSSASASLIEQOQSAQLOAPQPT-----SDSVKKEPVLAQPAVNSGTAASSTSLV 1561
QY 845 QLAENSVPVP-VGAOKTALRISOSTFNPNRPGOLGRLLNPNLPLDITLQSPGAGPF- 902
DB 1562 ALSAENTPATGYPDARTEAVPPASSFS--VPGQTA-----VIAAATSSACPV 1608
QY 903 -----PPIRNSPYSVLPQD-----GMGNOGMIGNOGNLGNSSTG---MIGNSASRP 947
DB 1609 VETNSPPIASTSTSIYAPGSAEAAARGVTSGSSVFAQPPAASSSSAFNQLNTMTAT-- 1666
QY 948 TMSGEMAPQSSAVRYTCAATTSAMNRPVQGMIRNPASIPMRPS-----OPG 997
DB 1667 -----APSATVFQGVAASTASLEFGQQTGSTASTAATAATPOVSSSGFSFSPAGTTAPG 1719
QY 998 Q-ROTLQSQVMNIGPSLE-----MNMGGPOYSQOQA--PNNQAPMWSLELPIDAQSFAS 1050
DB 1720 VFGQTTFGQASVFGQSSASASVFSFSGPQGFSSVPAGFGPASTPTSTSGSVFGAASSTS 1779
QY 1051 QNRQ-FPGSSPDDLCPHAPAESPSDEGALLDQLYL-ALBNFDGLEIDRALGIPELVSQ 1108
DB 1780 SSSSFSFGQS-----SPNTGGGLFGQSNAPAFQSGPFGGSGVFGTSAATY 1827
QY 1109 SQAV-----DPROFSSQDSNIMLEQKAP---VFPOQVYSAQMAQAGSYSPMODPNTFHT 1158
DB 1828 TAATSGSFSCQASGFGSSNGSVFGQAASTGTVFGQSSSSSGSVSGNTGCGGFGFS 1887
QY 1159 -MGORPSYATLRMQPRGLPRTGLVQNPQNLRLQLQLHQAQONRPLMN-----Q 1209
DB 1888 GLGGRS-----QDAANKPFSASGFGSTA 1914
QY 1210 ISNVSNVNLTRPGVPTQADINQMLAQROREILNHLROROMHQOQVOQRTLMRGQ 1269
DB 1915 TSMTS--NLFGNSGAKTFEGFASSTFGE-----QKPGTFFSGG 1951
QY 1270 LNMTPSVNASSGMPATNSNRIPOANAQOPPPPNVIGISQDPDGFATTPQSLMSPR 1329
DB 1952 ----GSVASQGGFSSPNTGFGGAAPVFGSPTEFGS---PGFGVPAFGSA----- 1997
QY 1330 MAHTQSPMQOQANPYAQAPSDINGMAQMGNSMFSQOSP-----HFGQANTSM 1383
DB 1998 -----PAFTIP-----LSTGG-KVFGECTAAASGSGGFGSSMT-- 2032
QY 1384 YSNMNINVSMAATNTGSMSSMNQMT-QQIS-MTSVTSVSTGLSSMG 1428
DB 2033 -----TSFGTLASQNAPTFGLSLQQTSGFGTQSSGFGFG 2067

RESULT 21
SIML_MOUSE
ID SIML_MOUSE STANDARD; PRT; 765 AA.
AC 061045; P70183;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Single-minded homolog 1 (Sim1).
GN SIM1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SWISS WEBSTER;
RX MEDLINE=97020303; PubMed=8812055;
RA Fan C.-M., Kuwana E., Bulfone A., Fletcher C.F., Copeland N.G.,
RA Jenkins N.A., Crews S., Martinez S., Puellas L., Rubenstein J.L.,
RA Tessier-Lavigne M.;
RT "Expression patterns of two murine homologs of Drosophila
RT single-minded suggest possible roles in embryonic patterning and in
RT the pathogenesis of Down syndrome.";
RL Mol. Cell. Neurosci. 7:1-16(1996).

```

```

RN [2]
RP ERRATUM.
RX MEDLINE=97029422; PubMed=8875433;
RA Fan C.-M., Kuwana E., Bulfone A., Fletcher C.F., Copeland N.G.,
RA Jenkins N.A., Crews S., Martinez S., Puellas L., Rubenstein J.L.,
RA Tessier-Lavigne M.;
RL Mol. Cell. Neurosci. 7:519-519(1996).
RN [3]
RP REVISIONS TO C-TERMINUS.
RX MEDLINE=97343329; PubMed=9199934;
RA Fan C.-M.;
RL Unpublished results, cited by:
RL Cirast R., Scott H.S., Chen H., Kudoh J., Rossier C., Minoshima S.,
RL Wang Y., Shimizu N., Antonarakis S.E.;
RL Genome Res. 7:615-624(1997).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RX MEDLINE=96413339; PubMed=8927054;
RA Ena M., Morita M., Ikawa S., Tanaka M., Matsuda Y., Gotoh O.,
RA Saijoh Y., Fujii H., Hamada H., Kikuchi Y., Fujii-Kuriyama Y.;
RT "Two new members of the murine Sim gene family are transcriptional
RT repressors and show different expression patterns during mouse
RT embryogenesis.";
RL Mol. Cell. Biol. 16:5865-5875(1996).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SV;
RA Hosoya T.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SUBUNIT.
RX MEDLINE=97172525; PubMed=9020169;
RA Probst M.R., Fan C.M., Tessier-Lavigne M., Hankinson O.;
RT "Two murine homologs of the Drosophila single-minded protein that
RT interact with the mouse aryl hydrocarbon receptor nuclear
RT translocator protein.";
RL J. Biol. Chem. 272:4451-4457(1997).
CC -|- FUNCTION: TRANSCRIPTIONAL FACTOR THAT MAY HAVE PLEIOTROPIC EFFECTS
CC DURING EMBRYOGENESIS AND IN THE ADULT.
CC -|- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER
CC BHLH PROTEIN. HETERODIMER OF SIM1 AND ARNT.
CC -|- TISSUE SPECIFICITY: DETECTED IN LONG, SKELETAL MUSCLE AND KIDNEY.
CC DURING FETAL DEVELOPMENT IT IS FOUND IN THE CNS, DEVELOPING
CC KIDNEY, MESODERMAL AND ENDODERMAL TISSUES, INCLUDING DEVELOPING
CC SOMITES, MESONEPHRIC DUCT, AND FOREGUT.
CC -|- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS.
CC -|- SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.
CC -|- SIMILARITY: CONTAINS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: U40575; AAA91201.1; ALT_SEQ.
DR EMBL: D79209; BAA11467.1; -.
DR EMBL: AB013491; BAA28270.1; -.
DR EMBL: AB013484; BAA28270.1; JOINED.
DR EMBL: AB013485; BAA28270.1; JOINED.
DR EMBL: AB013486; BAA28270.1; JOINED.
DR EMBL: AB013487; BAA28270.1; JOINED.
DR EMBL: AB013488; BAA28270.1; JOINED.
DR EMBL: AB013489; BAA28270.1; JOINED.
DR EMBL: AB013490; BAA28270.1; JOINED.
DR MGD: MGI:98306; Sim1.
DR InterPro: IPR003015; HLH_Myc.
DR InterPro: IPR001092; HLH_dim.
DR InterPro: IPR001067; NucleinsLocator.

```



DR InterPro: IPR001610; PAC.  
 DR InterPro: IPR000014; PAS.  
 DR Pfam: PF00785; PAC; 1.  
 DR Pfam: PF00983; PAS; 2.  
 DR PRINTS: PR00785; NCTRNLOCATR.  
 DR SMART: SM00353; HLH; 1.  
 DR SMART: SM00086; PAC; 1.  
 DR SMART: SM00091; PAS; 2.  
 DR PROSITE: PS00036; HELIX\_LOOP\_HELIX; 1.  
 DR PROSITE: PS00112; PAS; 2.  
 DR Developmental protein; Neurogenesis; Nuclear protein; Repeat;  
 KW Transcription regulation; DNA-binding.  
 FT DNA\_BIND 1 13 BASIC DOMAIN.  
 FT DOMAIN 14 54 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).  
 FT DOMAIN 77 147 PAS 1.  
 FT DOMAIN 218 288 PAS 2.  
 FT CONFLICT 133 133 H->L (IN REF. 1).  
 FT CONFLICT 176 176 MISSING (IN REF. 1).  
 FT CONFLICT 322 322 P->R (IN REF. 1).  
 FT CONFLICT 480 480 A->P (IN REF. 1).  
 FT CONFLICT 537 537 D->S (IN REF. 1).  
 SO SEQUENCE 765 AA; 85540 MW; B1A7FDAB578CD17 CRC64;

Query Match 3.5%; Score 266; DB 1; Length 765;  
 Best local similarity 20.2%; Pred. No. 2.8e-05;  
 Matches 162; Conservative 110; Mismatches 302; Indels 228; Gaps 33;

QY 32 RTEKRNREQENKYLEELAELEAFANFNIDNENFRP-----DKCAILEKTVQAIR 81  
 DB 6 KNAATREKENSEFEYELAKLL-----PLPSAIRSQDKASITILTSYKLM 52

QY 82 QIKDEKAAANIDEVQSDVSSTGQVIDKALDPRMELADGFFPVNLEGNNVFVSE 141  
 DB 53 RVFPEGLGCEAMGHTSRSPSPDNVGR-----ELGSHLLQTLTGDFPVVAPGKIMYISE 106

QY 142 NTVQRLRYNOELRNKSVYSILHVDHTEFVKNLPRKSIIVNGSGWSGEPFRNSHTFNCR 201  
 DB 107 TASYHLGISOVELTGNSTIYEYIHRADHDHETAVL----- 140

QY 202 MLVRLPDPSEEGHNDQEN--QKYE-----TMOCFRANVSQPKIKERGEDLQSC---- 248  
 DB 141 -----TAHQPYSHFEVQYEIERSEFRLMKCYLAKNAGLTCGGYKVIHCGYL 189

QY 249 -----LTCVARVPMKRPVLP-----SESFTRDLOQKI 280  
 DB 190 KIRQYSLDMSPPDGCYQNVGLVAVGHSLP-----PSAVTEIKLHSNFMFRASIDMKL 242

QY 281 TSLDTSTMRAMKPGME--DIVRRCIOKFHAOGESESVYAKRHHHEVLRQGLAFSQTIR 338  
 DB 243 IFLDS---RVAELTGERODDLEKTL--YHNHV--GCDTFHLKCAHHLLLVKQVYTKYR 296

QY 339 FSLSDGLVAAGTQSKLIRSQTTNEPOLVLSLHMLHREQNCVMMPDILTGQTMGRPLNDI 398  
 DB 297 FLAKGGGVMWQSYATIVHNSRSSRPHCIYSV-----NYVLDTPEYKGLDS--LDQI 347

QY 399 SSNSNAHOLCSGNCQDWTLSNINFPNGKPEQMGMPGSGGNNHVSQMATI-- 457  
 DB 348 SASKTFEYTSSTP-----TISDN-----RKGAKSRLSSSKSKSRKSP 386

QY 458 -PGGSNY--ALKMNSPSSQSPGNPCOPTS---MLSPRHMSGVAGSPRIIPSPSPAS 512  
 DB 387 FYQISGFITERESDHDSOMGSPPLTDIASPOLDBER---PGSQHELSCAVRQPPDRSS 443

QY 513 L-----HSPYGVCSSTGNSHSYTNSSLNA-----LQALSRGHVSLGSSLASPD 557  
 DB 444 LCYGFALDHSRL-----VEDRHFHTQACGRCCEAGRYFLGAPPGRDPMWMSRAALPLT 498

QY 558 KMGNTLQNSFVMMNPPLSLMGLSKDCFGLYGEPSEGTGAESSCHGEQKET-----N 613  
 DB 499 KASPSREAVENSMPIHSIRIHRIR--GHWDEDSVWSSPPGSGSESGDRYRTEYOQN 555

QY 614 DNLPPAVASSERADGQSLHDSKQTKILLQLLTTSKSDQNEPPLASLSLDTNKKDSGSLP 673

DB 556 SPHEPSKIETLIRATQOMKKEENR---LQLKAPPDQ-----ASIN 595  
 QY 674 GSGTHGTSLSKAKH-----KILHRLDSSSPVDLAKITLAFATKDS-----QESSST 722  
 DB 596 GAGKKHSLCFFANYQOPPTPGYCHSSALASTSPCDHIQ---QREGKMLSPHENDYNSPT 652

QY 723 A-----PGSEVITKOEPPSPK 738  
 DB 653 ALSRISSPSSDRITKSSLLIAK 674

RESULT 22  
 BUN2\_DROME STANDARD; PRT; 1211 AA.  
 AC Q24523; Q9VK78; Q9VK79;  
 DT 01-MAR-2002 (Rel. 41, Created)  
 DT 01-MAR-2002 (Rel. 41, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Bunched protein, class 2/class 3 isoforms (Shortsighted protein).  
 GN BUN OR SHS OR CG5461.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A. (CLASS 2 ISOFORM), SUBCELLULAR LOCATION, AND  
 RP FUNCTION.  
 RC TISSUE=Eye-antennal disk;  
 RX MEDLINE=96038094; PubMed=755710;  
 RX Treisman J.E., Lai Z.-C., Rubin G.M.;  
 RT "Shortsighted acts in the decapentaplegic pathway in Drosophila eye  
 RT development and has homology to a mouse TGF-beta-responsive gene";  
 RL Development 121:2835-2845(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A. (CLASS 2 AND CLASS 3 ISOFORMS).  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RX Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RX Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 RX George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RX Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RX Brandon R.C., Rogers Y.-H.C., Blazer R.G., Chame M., Pfeiffer B.D.,  
 RX Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RX Abriti J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RX Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RX Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,  
 RX Borokova D., Botchan M.R., Bouck J., Brostein P., Brotler P.,  
 RX Burtis K.C., Busan D.A., Butler H., Cadien E., Center A., Chandra I.,  
 RX Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RX de Rubeis B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RX Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RX Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,  
 RX Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RX Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RX Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RX Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RX Jatali M., Kalush F., Karpen G.H., Ke Z., Kennison J., Ketchum K.A.,  
 RX Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RX Lasco P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RX Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RX Merulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RX Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RX Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pauley J.M.,  
 RX Palazozo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RX Reibert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RX Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RX Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RX Svitzkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RX Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RX Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RX Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,



DB 898 SYAAPQALPTLQASAPSTIADPQALMPQOQOQOHOE-----EQO 939

OY 1371 SPHFQOQ-----AN-TSMYSNNMNVSMATNGSMXNMTGOISMT 1414

DB 940 QQPQOQOQOPLPPANIASASANSNLT-----TNTNVATGEATNTALTLT 985

RESULT 23

ARN2\_HUMAN STANDARD: PRT: 706 AA.

ID 09H8Z2: 015024; (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Aryl hydrocarbon receptor nuclear translocator 2 (ARNT protein 2).

GN ARNT2 OR KIAA0307.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI\_taxid=9606;

RN [1]

RP SEQUENCE FROM N.A.

RA Barrow L.L., Wines M.E., Romitti P.A., Holdener B.C., Murray J.C.;

RT "Aryl-hydrocarbon receptor nuclear translocator 2 (ARNT2): structure,

RT gene mapping, and candidate evaluation for human orofacial clefts.";

RL submitted (SEP-1999) to the EMBL/Genbank/DBJ databases.

RM [2]

RN SEQUENCE FROM N.A.

RC TISSUE=Brain;

RX MEDLINE=9734984; PubMed=9205841;

RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,

RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;

RT "Prediction of the coding sequences of unidentified human genes. VII.

RT The complete sequences of 100 new cDNA clones from brain which can

RT code for large proteins in vitro.";

RL DNA Res. 4:141-150(1997).

CC -1- FUNCTION: SPECIFICALLY RECOGNIZES THE XENOBIOTIC RESPONSE ELEMENT

CC (XRE).

CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER

CC HIGH PROTEIN. HETERODIMER WITH THE ARYL HYDROCARBON RECEPTOR (AHR)

CC OR THE SIM1 PROTEIN (By similarity).

CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).

CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF

CC TRANSCRIPTION FACTORS. HIGHEST TO THE ARNT PROTEINS.

CC -1- SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.

CC -1- SIMILARITY: CONTAINS 1 PAS-ASSOCIATED C-TERMINAL (PAS) DOMAIN.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL: AF185610; AAG15310.1; -

DR EMBL: AF185593; AAG15310.1; JOINED.

DR EMBL: AF185594; AAG15310.1; JOINED.

DR EMBL: AF185595; AAG15310.1; JOINED.

DR EMBL: AF185596; AAG15310.1; JOINED.

DR EMBL: AF185597; AAG15310.1; JOINED.

DR EMBL: AF185598; AAG15310.1; JOINED.

DR EMBL: AF185599; AAG15310.1; JOINED.

DR EMBL: AF185600; AAG15310.1; JOINED.

DR EMBL: AF185601; AAG15310.1; JOINED.

DR EMBL: AF185602; AAG15310.1; JOINED.

DR EMBL: AF185603; AAG15310.1; JOINED.

DR EMBL: AF185604; AAG15310.1; JOINED.

DR EMBL: AF185605; AAG15310.1; JOINED.

DR EMBL: AF185606; AAG15310.1; JOINED.

DR EMBL: AF185607; AAG15310.1; JOINED.

DR EMBL: AF185608; AAG15310.1; JOINED.

DR EMBL: AF185609; AAG15310.1; JOINED.

DR EMBL: AB002305; BAA20766.1; -

DR MIM: 606036; -

DR InterPro: IPR001092; HLH\_dlm.

DR InterPro: IPR001067; Nucleinslocator.

DR InterPro: IPR001610; PAC.

DR InterPro: IPR000014; PAS.

DR Pfam: PF00010; HLH; 1.

DR Pfam: PF00989; PAS; 2.

DR PRINTS: PR00785; NCTRNSLOCATR.

DR SMART: SM00353; HLH; 1.

DR SMART: SM00086; PAC; 1.

DR SMART: SM00091; PAS; 2.

DR PROSITE: PS00038; HELIX\_LOOP\_HELIX; 1.

DR PROSITE: PS0112; PAS; 2.

KW DNA-binding; Nuclear protein; Transcription regulation; Repeat.

FT DNA\_BIND 53 65

FT DOMAIN 66 106 BASIC DOMAIN.

FT DOMAIN 123 198 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).

FT DOMAIN 312 382 PAS 1.

FT DOMAIN 387 430 PAS 2.

FT DOMAIN 62 65 PAC.

FT DOMAIN 429 434 POLY-ARG.

FT DOMAIN 541 544 POLY-SER.

FT DOMAIN 64 64 R -> K (IN REF. 2).

FT CONFLICT 668 668 G -> S (IN REF. 2).

SQ SEQUENCE 706 AA; 77612 MW; 2187A3C603EC7D39 CRC64;

Query Match 3.4%; Score 262; DB 1; Length 706;

Best Local Similarity 21.9%; Pred. No. 3.9e-05;

Matches 163; Conservative 106; Mismatches 285; Indels 190; Gaps 37;

OY 11 PSR-AETKRKKECPQOLSPKPR-----NTEKRNQDENKYTEELAEILFANNDIDN 62

DB 29 PARGGKRRSGMDPDEDEGSPKFSRENHSEIERRRRRKMQYITLSDM-----PTCSA 84

OY 63 FNFKPKCAILKETVKQJROIKECEKAAANIDEVQKSDVSTGGVTDKDALGPMLEA 122

DB 85 LARKDKLTILMAVSHHKSNR-----GTGN-----KTDGAYKRSFTLEDKILILEA 134

OY 123 LDGFFPVNLE-GNVVFESENVTOYLRYNOBELMKNKSVYSLHVDHTEPKNLLPKSIV 181

DB 135 ADGFLFVAAEGRVIVYSDSVTPVLNQPSMFCSTLYEYVHPD-----VEKLEQDCT 190

OY 182 NGGSMG-----EPPRRNS-----HFNCRMLVKKPLP-----DSEEE 213

DB 191 SENSMTGRILDLKTGTVYKKEGQSSMRMCKGSRSFICRMCGNAPLDLPLNRTTMRK 250

OY 214 GHND-----QEAHOKYETMOC-----FAVSQPKSITKEGEDLOSLICVARVPK 259

DB 251 RFRNGLGVYKGEAOYAVVHCTGYIKAMPYPAGMTIPEDDADVGOSKYLVAIG-RLOYT 309

OY 260 ERPV-----LPSSSEFTTRQDLOGKITSLDSTWRAAMKPGW--EDLVRCIOKF-HAQ 310

DB 310 SSPVCMDNKMSVPTPEFLSRHNSDGIITFVDP--RCISVIGYQPDLLGKDIIEFCRP- 365

OY 311 HEGEVSYAKKHHHYLR-QGLAFSQIYRFSISDGTILVAADTKSLINSQTTNEQLVI- 368

DB 366 ---EDQSHLRSPQOVVVKLGQVLSVMYRFRKKNEMWLITSSFTFQNPYSDELEYIC 422

OY 369 ---SLHMLHREQ-----NVCVMNPDLTGQTMGKPLNPIS-----NSP 403

DB 423 TINTVTKQLOQOQDALELVHQRGLSSYDLSQYVPVRLPAGVHACKSVKAKAITSQEDDP 482

OY 404 AHQALCSGNPQODMTLSSNINPLNGPKQMG--MPMGREG--SGGNHVSQMATTPO 459

DB 483 RFAEFACIASASEKKMMSASAGAQTQIYSOGSPFSGSHKARSSVYHVPGVN----- 537

OY 460 GSNVALKMNPSQSSPGNPPGPTSMLSPRHRMSGVAGSRIP---PSQFSPAGLSHS 515

DB 538 -----DIQSSSTGNNMSQISKRLNOSQVAMTG--SRPFPQOQIIPSSSSTQOS--S 585

OY 516 PVGVCSSTGNSHSYTNSLSLNLQALSEGHGVSGLASSLASP--DLKMGNIQNSPVNMNPP 573

```

Db      586 PFGI-----GSHYYPADP-----SSYSPSSPATSPSCSNAYSLANRP-- 626
OY      574 LSKMSLSDSDKDFGLY-GEPSEGTGQAESSCHPGEOKEKENDPMLPAVSSERADGOSRL 632
Db      627 -----GFAESQSSSQGFCGRSE-VMSQWQSQ-HHGQO-----SGEQHS 663
OY      633 HDSKGQTKLLQLLTYSKSDMEPSP 656
Db      664 HQQPGTEVFQ-----DMLPMF 680

RESULT 24
ID      HIFA_MOUSE STANDARD; PRT: 822 AA.
HIFA_MOUSE
AC      061221; 061665; 061664; 008993; 008741;
DT      01-NOV-1997 (Rel. 35, Created)
DT      15-DEC-1998 (Rel. 37, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Hypoxia-inducible factor 1 alpha (HIF-1 alpha) (ARNT interacting
GN      HIF1A.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX      NCBI_Taxid=10090;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=C57BL/6; TISSUE=Hepatocytes;
RX      MEDLINE=96355491; PubMed=8702901;
RA      Li H., Ko H.P., Whitlock J.P. Jr.;
RT      Induction of phosphoglycerate kinase 1 gene expression by hypoxia.
RT      J. Biol. Chem. 271:21262-21267(1996).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=129/SVJ;
RX      MEDLINE=98034461; PubMed=9368100;
RA      Luo G., Gu Y.-Z., Jain S., Chan W.K., Carr K.M., Hogenesch J.B.,
RA      Bradfield C.A.;
RT      Molecular characterization of the murine Hif-1 alpha locus.
RT      Gene Expr. 6:287-299(1997).
RN      [3]
RP      SEQUENCE FROM N.A.
RC      STRAIN=129/SV;
RX      MEDLINE=97354184; PubMed=9210478;
RA      Wenger R.H., Rolfs A., Kvietikova I., Spielmann P., Zimmermann D.R.,
RA      Gassmann M.;
RT      "The mouse gene for hypoxia-inducible factor-1alpha. Genomic
RT      organization, expression and characterization of an alternative first
RT      exon and 5' flanking sequence."
RT      Eur. J. Biochem. 246:155-165(1997).
RN      [4]
RP      SEQUENCE OF 13-822 FROM N.A.
RC      TISSUE=Hepatocytes;
RX      MEDLINE=96254028; PubMed=8660378;
RA      Wenger R.H., Rolfs A., Marti H.H., Guenet J.-L., Gassmann M.;
RT      "Nucleotide sequence, chromosomal assignment and mRNA expression of
RT      mouse hypoxia-inducible factor-1 alpha."
RT      Biochem. Biophys. Res. Commun. 223:54-59(1996).
RN      [5]
RP      SEQUENCE OF 22-85 FROM N.A.
RC      TISSUE=Hepatocytes;
RA      O'Rourke J.F.;
RL      Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
CC      -1- FUNCTION: INVOLVED IN THE INDUCTION OF OXYGEN REGULATED GENES.
CC      SPECIFICALLY RECOGNIZES AN 8 BP HYPOXIA RESPONSE ELEMENT (HRE).
CC      -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER
CC      BHLH PROTEIN. HETEROIMER OF AN ALPHA AND A BETA (ARNT) SUBUNITS.
CC      THE ALPHA SUBUNIT IS UNIQUE TO HIF-1 WHEREAS HIF-1 BETA (ARNT) CAN
CC      DIMERIZE WITH OTHER BHLH-PAS PROTEINS. INTERACTS WITH HSP90 (BY
CC      SIMILARITY).
CC      -1- SUBCELLULAR LOCATION: Nuclear (Potential).

```

```

CC      -1- TISSUE SPECIFICITY: UBIQUITOUS.
CC      -1- DOMAIN: BOTH HYPOXIA RESPONSIVENESS AND TRANSACTIVATION CAPABILITY
CC      RESIDE WITHIN THE C-TERMINAL PART.
CC      -1- PTM: REQUIRES PHOSPHORYLATION FOR DNA-BINDING.
CC      -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC      TRANSCRIPTION FACTORS.
CC      -1- SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.
CC      -1- SIMILARITY: CONTAINS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC
DR      EMBL: U59496; AAC52730.1; -.
DR      EMBL: AF003695; AAC53455.1; -.
DR      EMBL: Y09085; CAA70306.1; -.
DR      EMBL: Y13656; CAA70306.1; JOINED.
DR      EMBL: Y09085; CAA70305.1; -.
DR      EMBL: AF004155; AAC53461.1; -.
DR      EMBL: AF004141; AAC53461.1; JOINED.
DR      EMBL: AF004142; AAC53461.1; JOINED.
DR      EMBL: AF004143; AAC53461.1; JOINED.
DR      EMBL: AF004144; AAC53461.1; JOINED.
DR      EMBL: AF004145; AAC53461.1; JOINED.
DR      EMBL: AF004146; AAC53461.1; JOINED.
DR      EMBL: AF004147; AAC53461.1; JOINED.
DR      EMBL: AF004148; AAC53461.1; JOINED.
DR      EMBL: AF004149; AAC53461.1; JOINED.
DR      EMBL: AF004150; AAC53461.1; JOINED.
DR      EMBL: AF004151; AAC53461.1; JOINED.
DR      EMBL: AF004152; AAC53461.1; JOINED.
DR      EMBL: AF004153; AAC53461.1; JOINED.
DR      EMBL: AF004154; AAC53461.1; JOINED.
DR      EMBL: X95580; CAA64833.1; -.
DR      EMBL: X95002; CAA64458.1; -.
DR      MGD: MGI:106918; H1fla.
DR      InterPro: IPR003015; HLH_Myc.
DR      InterPro: IPR001092; HLH_dlm.
DR      InterPro: IPR001321; HypoxIndf1A.
DR      InterPro: IPR001610; PAC.
DR      InterPro: IPR000014; PAS.
DR      Pfam: PF00785; PAC; 1.
DR      Pfam: PF00989; PAS; 2.
DR      PRINTS: PR01080; HYPOXIAFLA.
DR      SMART: SM00353; HLH; 1.
DR      SMART: SM00086; PAC; 1.
DR      SMART: SM00091; PAS; 2.
DR      PROSITE: PS00038; HELIX_LOOP_HELIX; 1.
DR      PROSITE: PS50112; PAS; 2.
KW      Repeat; DNA-binding; Nuclear protein; Transcription regulation;
KW      Activator; Phosphorylation.
FT      DNA_BIND 17 30
FT      DOMAIN 31 71 BASIC DOMAIN.
FT      DOMAIN 80 155 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
FT      DOMAIN 228 298 PAS 1.
FT      DOMAIN 302 345 PAS 2.
FT      CONFLICT 31 31 S->T (IN REF. 1).
FT      CONFLICT 128 128 A->T (IN REF. 2 AND 3).
FT      CONFLICT 351 351 I->L (IN REF. 1).
FT      CONFLICT 511 511 E->L ERLQENVNTPNFSQ (IN REF. 2).
FT      CONFLICT 686 686 K->N (IN REF. 3 AND 4).
FT      CONFLICT 785 785 E->V (IN REF. 4).
SQ      SEQUENCE 822 AA; 91858 MW; E717ADCEAC9D795 CRC64;

```

Query Match 3.4%; Score 262; DB 1; Length 822;  
 Best Local Similarity 19.6%; Pred. No. 4.7e-05;  
 Matches 178; Conservative 132; Mismatches 329; Indels 268; Gaps 38;

QY	5	GENTSDPSAEIRKKECPDOLGSPKNTTEKRNNEQCNKIIEELAEILFANFNIDINFN	64
Db	6	GENEKKKMSSERRKK-----SRDAASRRKSEFVELAHOJLPLPHVSSHL-	54
QY	65	FRPDCALILEKVEVAKIROIKERQEKAAANIDEVOKSDVSTGGQYIDALGPMLEALD	124
Db	55	---DKASVWRLLISLVRKRLLDAGGIDSEDEM-----KAQMDCYLRALD	97
QY	125	GEFFVNVLEGNVVESENAVTOYLRYNOBELMKNKSVYSILHVGDTHEFYKNLLPKRSYNGG	184
Db	98	GFVAVLTGDCGMVYISDVNKKYMGCLTQELAGHSVFDETHPCDHHEKREMLTHR--NG-	15
QY	165	SKMSGPPRRNRNHTFCR---MLKVPRLDSEEGHDNOEAOK-----YET----	222
Db	154	-----PRKGEKELTQTSFELRMKCTLTLSGRMTNIKSATWVKLHCTGHIHVYDTFNSNQ	207
QY	227	MOCFAVSOPKSIKEGEEDLOSCILCAVRVVRMKERPVLP--SESEPTTRQDLOGKITSLDT	289
Db	208	PQCGKRRP-----MTCLVILCPRIHPNSNIEIPLDLSKTFPLSHSIDMKFSTCDE	25
QY	286	S--TMRAAMKPEMEDLVRRCIQKFNAAHGBESVSYAKRHHNEVLRQGLAFASQIYRFSISDG	344
Db	258	RITELMGYEP--EELLGSIYEY--HALDS-DHLTKTHHDMFKGQVTTQGYHMLAKRG	312
QY	345	TLVVAQTSKSLIRSGTTHEPQVLSLH-----MLHREONCVAMP--DLTGCTMG	392
Db	313	GYVWVEVQATVYINTKNSQPOCIYCVNVVSGIIIOHDLIFSLQOTESYLKFEVESDDMKMT	372
QY	393	KPLNPIS-----NSPAHQALCSGNRG-----ODMTLSS	422
Db	373	QLTFKVESEDTSCLFDKLAKKEPDALTILAPAAGDTIISLDFGSDTTEDDQULEDPVLYN	433
QY	422	NINFPINGPKBOMGPMG-----REFGSGGN--HYSGQAATTPQ	455
Db	433	DVMPSSNEKININILAMSPLEPSETPKPLRSSADPALNOEVALKLESSPESLGLSTMPQ	492
QY	460	-----GSNVALKMNSSQSSP-----GNP--GQ	483
Db	493	IQDDPASPSDSTGROSSPEPNSFSEYCFDVSDDMVNVFKLELYEKLEPAEDTEAKNPESTQ	552
QY	482	PT---SMLSPR-----HRMSPGVAQSPRIIPPSQFSPAG---SLHSPGVCS	521
Db	553	DTDIDLEKLAVIYIMDDDFQLRSPDQLSPLESNPS--PPSMSTYTGQOQYOLQKP--TIT	609
QY	522	STGNSHSTYNSLNAQALSLSGHGVSLGSSLASPDLKMGILONSPPVMMNPPLSKMGLD	581
Db	610	ATATTATTATDESKEITKDNKE-----DIKI--LIASPSSTQ-----	643
QY	582	SKDQFGLGSESECTGOAESCHPGEDEKTEINDNLPNAYSSPADQOSRLHDSKGTKL	641
Db	644	-----VQETTT--AKASAVSGTHSRASPSDRAGKRVIEQFD--KAPRSLKLSATL	691
QY	642	LQLLTFRSDQMEPPLASSLSIDTK--DSTSL--DGSGSTHGTSLKEKHKILHRLLOD	696
Db	692	NQRTVVEEELNPTITASQNAQRKRMHDSLQAAIGT-----LLQO	736
QY	697	SSSPVDLAKITAEATGKDLQSSSTAPGSEFVITIKQEPVSPKKENALLRYLDDKDDTKD	756
Db	737	PGDCAPITMSMKRWKGFISEQNGTBOKTIILL-----PSDLACRLLAGSDNES----	786
QY	757	IGLPEIT	763
Db	787	GLPOLT	792
RESULT	25	ARN2_MOUSE	
AC	Q61324;	STANDARD;	PRT; 712 AA.
DT	15-DEC-1998 (Rel. 37, Created)		
DT	15-DEC-1998 (Rel. 37, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Aryl hydrocarbon receptor nuclear translocator 2 (ARNT protein 2).		

GN ARNT2.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID:10090.  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Embryo;  
RX MEDLINE=96239527; PubMed=8657146;  
RA Hirose K., Morita M., Enna M., Mimura J., Hamada H., Fujii H.,  
RA Saito Y., Gotch O., Sogawa K., Fujii-Kuriyama Y.,  
RT "cDNA cloning and tissue-specific expression of a novel basic  
RT helix-loop-helix/PAS factor (ArntL) with close sequence similarity  
RT to the aryl hydrocarbon receptor nuclear translocator (Arnt).";  
RL Mol. Cell. Biol. 16:1706-1713(1996).  
CC - FUNCTION: SPECIFICALLY RECOGNIZES THE XENOBIOTIC RESPONSE ELEMENT  
CC (XRE).  
CC - SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER  
CC BHLH PROTEIN. HETERODIMER WITH THE ARYL HYDROCARBON RECEPTOR (AHR)  
CC OR THE SIM1 PROTEIN.  
CC - SUBCELLULAR LOCATION: Nuclear (Potential).  
CC - TISSUE SPECIFICITY: RESTRICTED TO ADULT BRAIN AND KIDNEY.  
CC - SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF  
CC TRANSCRIPTION FACTORS. HIGHEST TO THE ARNT PROTEINS.  
CC - SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.  
CC - SIMILARITY: CONTAINS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.  
CC  
CC THE SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL: D63644; BAA09799.1; -.  
DR MGD: MGI:107188; Arnt2.  
DR InterPro: IPR003015; HLH\_Myc.  
DR InterPro: IPR001092; HLH\_dim.  
DR InterPro: IPR001067; Nuclnslocator.  
DR InterPro: IPR001610; PAC.  
DR InterPro: IPR000014; PAS.  
DR Pfam: PF00010; HLH; 1.  
DR Pfam: PF00989; PAS; 2.  
DR PRINTS: PR00785; NCTRNSLOCATR.  
DR SMART: SM00353; HLH; 1.  
DR SMART: SM00086; PAC; 1.  
DR SMART: SM00091; PAS; 2.  
DR PROSITE: PS00038; HELIX\_LOOP\_HELIX; 1.  
DR PROSITE: PS50112; PAS; 2.  
DR DNA-binding; Nuclear protein; Transcription regulation; Repat.  
KW DNA-BINDING; Nuclear protein; Transcription regulation; Repat.  
FT FT 64 76 BASIC DOMAIN.  
FT FT 77 117 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).  
FT FT 134 209 PAS 1.  
FT FT 323 393 PAS 2.  
FT FT 398 441 PAC.  
FT FT 73 76 POLY-ARG.  
FT FT 440 445 POLY-GLN.  
FT FT 552 555 POLY-SER.  
SQ SEQUENCE 712 AA; 77884 MW; FACCDD2EEBB7F18DE CRC64;

Query Match 3.4%; Score 260.5; DB 1; Length 712;  
Best Local Similarity 21.8%; Pred. No. 4,6e-05;  
Matches 162; Conservative 109; Mismatches 279; Indels 193; Gaps 36;

QY 11 PSR-AETRRKRCPCDOLGSPSKR-----NTEKRNROENKYIEBELAELIFANFNIDN 62  
DB 40 PARGGKRRSGMDFDEDEGSPSKFRSHNSFEIERRRRKKMQLYIELSDMV-----PTCSA 95  
QY 63 ENFKDKCAILKEIYKQIRQIKEDKKAANAANIDEVOKSDVSSTGCGVTDKALGPMMLEA 122  
DB 96 IARRKDKLTLIRMAVSHHKSMMR-----GTGN-----KSTDGAVYKPSFLTEDELKILLIEA 145

```

0Y 123 LDGFFPVNLE-GANVVPSEVNTOTYLRNDEBLMNKSVSYLHVGHDTEPEYKMLPKRSIV 161
Db 146 ADGFELEVAALBTGRATVYSDVSTVLNQPQSEWEGSYLREQVHDD---VEKIREDLCT 201
0Y 182 NGGSMWG-----EPRRNS-----HTFNCRLMYKPLP-----DSEEE 213
Db 202 SENSTTGRILDLTKGTGYKKEGQSSMRMCKSRRSFICRRNGCARPLDLPLNRITTMK 261
0Y 214 GHDN-----QEAHOKYETMOC-----FAYSOPKSIKEEGEDLOSLCYARRVPMK 259
Db 262 RFRNGLGPKVKEGEAOAYAVHCTGYIKAMPAGMTIPEEDADVGSGSKYCHVAIG-RLOVT 320
0Y 260 ERPV-----LPSSSEFTTRDLOGKITSILDTSTMPRAAMPKW--EDLVRRCIOKE-HAQ 310
Db 321 SSPVCMDSMGSVLPPEFLSRINSOGIITFVDP---RCLSVIGTPODULCKDLIEFCHP- 376
0Y 311 HEGESVSAYAKRHHHEVLR-QGLAEISOIYRFESLSDGTILVAOATKSKLRSQTTNEPOLVI- 368
Db 377 ---EDOSHILRESFOOVYKLGQVLSVMYRFRTRKNREMLLITRTSFTQNPYSDEIEYVIC 433
0Y 369 ---SLHMLHREQ-----NVCVNPDLITGOTMKPLNPITS---NSP 403
Db 434 TINTVWKQLOQOAOAELEVHQRDGLSSYDLSQVPRVNPPLAGVHEAGKSYEKADATFSQERDP 493
0Y 404 AHOALCSGNPGQDMTILSNINEFPIINGPREQMG--MPMGREFG--SGGMNHYSG---MQAT 456
Db 494 RFAEMFAGIASSEKKMMSASASSSQOITYSQSGSFPPAGHSKAPASSSVYHVPGVNDQSS 553
0Y 457 TPQGSNVYALKMNSPQSSPGMNPQOPTSMLSPRHMSPGVAGSPRIDPPSOFSPAGSLASP 516
Db 554 SSTGONI-----SQISROLNOCQ-----YAMTSSRPFPQGOESPQTQSSA 592
0Y 517 VGVCSSTGNSHYTNSNLATLQALSEBGVSLGSSLSL--DLKGMQLQMSVMMNMPPL 574
Db 593 FGI-----GSHHPYPADP-----SSYPLSSPAASSPFGAAYPSLAKRTP--- 632
0Y 575 SKMGSLDSKDCDFGLY-GEPSGTTGQAESSCHPEQCEKTPDNPPLPAVSEERADGGQSLH 633
Db 633 ---GRAESGSGGGQFGQGRPSE-VMSQMSQ-HHQOQ-----SGEESH 670
0Y 634 DSKGOTKLLOLTTKSDQMEPSP 656
Db 671 QOPGQTEVEFO-----DMLPMP 686

```

RESULT	26
NPAA2_HUMAN	
ID	NPAA2_HUMAN
AC	Q99743; Q99629;
DT	15-DEC-1998 (Rel. 37, Created)
DT	15-DEC-1998 (Rel. 37, Last sequence update)
DT	16-OCT-2001 (Rel. 40, Last annotation update)
DE	Neuronal PAS domain protein 2 (Neuronal PAS2) (Member of PAS protein 4) (MOP4).
DE	NPAS2.
GN	NPAS2.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX	NCBI_TaxID=9606;
RN	(1)
RP	SEQUENCE FROM N.A.
RX	MEDLINE=97165088; PubMed=9012850;
RA	Zhou Y.-D., Barnard M., Tian H., Li X., Ring H.Z., Francke U.,
RA	Shelton J., Richardson J., Russell D.W., McKnight S.L.;
RT	"Molecular characterization of two mammalian bHLH-PAS domain proteins selectively expressed in the central nervous system.";
RL	Proc. Natl. Acad. Sci. U.S.A. 94:713-718(1997).
RN	(2)
RP	SEQUENCE OF 1-626 FROM N.A.
RX	MEDLINE=97236817; PubMed=9079689;
RA	Hogensesch J.B., Chan W.K., Jackiw V.H., Brown R.C., Gu Y.-Z.,
RA	Pray-Grant M., Perdev G.H., Bradfield C.A.;

```

RT "Characterization of a subset of the basic-helix-loop-helix-PAS
RT superfamily that interacts with components of the dioxin signaling
RT pathway."
RL J. Biol. Chem. 272:8581-8593(1997).
CC -1 SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER
CC BHLH PROTEIN. INTERACTS WITH HSP90.
CC -1 SUBCELLULAR LOCATION: Nuclear (Potentlall).
CC -1 SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS.
CC -1 SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.
CC -1 SIMILARITY: CONTAINS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL: U77970; AAC47250.1; -.
DR EMBL: U51625; AAC51211.1; -.
DR MIM: 603347; -.
DR InterPro: IPR003015; HLH_MYC.
DR InterPro: IPR001092; HLH_dim.
DR InterPro: IPR001610; PAS.
DR InterPro: IPR000014; PAS.
DR Pfam: PF00010; HLH; 1.
DR Pfam: PF00785; PAC; 1.
DR Pfam: PF00989; PAS; 2.
DR SMART: SM00353; HLH; 1.
DR SMART: SM00086; PAC; 1.
DR SMART: SM00091; PAS; 2.
DR PROSITE: PS00038; HELIX_LOOP_HELIX; 1.
DR PROSITE: PS50112; PAS; 2.
KW Repeat: DNA-binding; Nuclear protein; Transcription regulation.
FT DNA_BIND 10 22
FT DOMAIN 23 60
FT DOMAIN 82 152
FT DOMAIN 237 307
FT DOMAIN 311 354
FT CONFLICT 51 51 K -> E (IN REF. 2).
FT CONFLICT 164 164 S -> G (IN REF. 2).
FT CONFLICT 308 308 T -> K (IN REF. 2).
FT CONFLICT 471 471 L -> S (IN REF. 2).
SQ SEQUENCE 824 AA; 91759 MW; 249AAC4687B328A5 CRC64;

```

```

Query Match 3 43; Score 259; DB 1; Length 824;
Best Local Similarity 19.4%; Pred. No. 6.4e-05;
Matches 215; Conservative 154; Mismatches 399; Indels 338; Gaps 46;

QY 4 MGENTSDPSRAETRRKRCPCDPLGSPKRNKFERRREOENKYIEELAILFANFNIDINF 63
Db 1 MDEEDKDAKRRSRK-----SEKKRRDDPNVILIKLSMLGNTR----- 41

QY 64 NEKPDKCAILKETVYKQIRKIQEKKAAANIDVOKSDVSTGQGVIDKDALGPMMLBAL 123
Db 42 --KMDKTVLEKVFELQ--KNEYSAQTEICDIDQDMKPS---FLSNEEFTQMLBAL 93

QY 124 DGEFFVVALLEGVNVFVSENVTOYLRIKNOBELMKNKSVSILIHGDTHEFVKLLPKSYNG 183
Db 94 DGFILAVTTDGSIIIVYSDISPIRLPLGLHLEPSVDWDMNLNPLQGEHSEVYKILSHMLVTD 153

QY 164 GSWSEPPRRNSHT---FNCRLVLPRLDSEEGND-----NOAHOKYETMOC----F 230
Db 154 ---SPSPYLAUSDSEFYCHLRLGSLNMPKEPTPEYIKFVGNFNSYNNVSPSCNGFDN 210

QY 231 AVSOKSKIEEGEDQSGCLICVARVP---MKERVLPSS--ESFTRODLOGKITSPTS 286
Db 211 TISRRCRV-PLAKEY--CFIATVRLATQOLFEMKCIVDEPRLDEFTSRSLSEKFLFLD-- 265

QY 287 TMAAMKPGM---EDLVRCIOKFPNAQHEGESVSAYAKRHHEVLRQGLAFSFOIYNFSLSD 343

```

```

Db 266 -HRAPIIGYLFEEVLGTSGYDYHI-----DDLELLARCHOHLMQFGTGKSCCYREFLFG 320
OY 344 GTLVAAQTKSKIRISQOTINEPOLVISTLHLEQVNCVNPULLTQOTACKPLNPISNSP 403
Db 321 OOWMLQTHYYTYTHQWMSKPEFIVCTHSVSYADV-----E 359
OY 404 AHQALCSGNPGODMTLSSNINFPINGPKQMGMPGREG-----GSGGNHVSQMATTPQ 459
Db 360 RQOELALDEPSEALHSSALK-----DKGSSLEPQHAFALDVAGSLN-----403
OY 460 GSNVALKNMSPQSSPGMNPQPTSMLSPRHMSR--GVAGSPRIIPPSQSPAGSLHSPV 517
Db 404 -----TSHSPSASSRSHKSSHTAMSEPTPTPKLMAEASPPALPRSATLPQ---ELPV 454
OY 518 GVCSTGSHSTNSLNAQLALSEGHSVGLSSIASP---DLKNGNLQNSVYNNMPPL 574
Db 455 -----PGLSQ--AATMPAPLPSPSLCDLTQOLLPOVYLOSTPAPM 492
OY 575 SKMGSLDKDCFGLYGEPESEGTGQAESSCHPGEQKETNDPPLPRAVSEERAD---GOSR 631
Db 493 AQFSNQ-----FSMF-----QTIKQLEORTRLQANIMQDEE 526
OY 632 LHDSKGQFKLLQLLTTKSDQNEPPLASLSLDPNKSSTGSLPGSGSTHOTSLEKHKILH 691
Db 537 LHKIQEOLCLVODSNVQMFLOPA-VSLFSSTQRPFA-----Q 564
OY 692 RLUDSSSPVDLAKLTAEATGKDLQSESSSTAPGEVITIKOPV-----SPKKENALLRY 747
Db 565 QOLQORSAAYVQPOLGA---GPQLPGQISSAQVTQSHLRESSVISTOGPKPMRSSOLMQ 621
OY 748 -----LDKDDTKDIGLEPITPKLERLDSKTPASNTKLIAMKTEKEKSFEPGDP 799
Db 622 SSGRGSLLVEPFSATAPL---PSL-----NLTPPASTS-----QD 656
OY 800 GSELDNLEILLDQLNSQLPOLFPDTPGAPAGSVDKAIIINDLMQLTAENSPTVPVGAQ 859
Db 657 ASQCPSPDFSHDQLRL--LSQIPQMP--GSCDAR-----QPEVSEVRTGQ 702
OY 860 KTAIRISQSTNNRRPGQLRLPNQNLPLDTTLOSPTGAGFPPIRNSPVSVPQGM 919
Db 703 VKYAO-SQTVFQNP-----DAHPANSSAPRPVL-----730
OY 920 MGNQMGIGNQNLGNSGTMGINSASRPTMPSGEWAPQSSAVRYCAATTSAMNRPVQGG 979
Db 731 -----LMQAVLHPSFASQ-----745
OY 980 MIRNPASIPMRPSSQPGOROTLOSQVANNIGPSELEMMNGGPOYSQ-----QAPPNQ 1032
Db 746 ---PSPHQPAQARQQPPQHY-LQVQAPTSLSHSEQDSILLSTYSQQFGTLGYPQPPAQ 800
OY 1033 TAPNPESILPIDQASFASQ---RQP 1055
Db 801 ---PQPLRPRAVSSLSSESGLOQP 822

```

```

RESULT 27
HRX_HUMAN
ID HRX_HUMAN STANDARD: PRT: 3969 AA.
AC Q03164; Q14845; Q16364; Q13743; Q13744; Q90MA3;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Zinc finger protein HRX (MLL-1) (Trithorax-like protein).
GN MLL OR HRX OR ALL1 OR TRX1 OR HTRX.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9304667; Pubmed=1423624;
RA Tkachuk D.C., Kohler S., Cleary M.L.;

```

```

RT "Involvement of a homolog of Drosophila trithorax by 11q23
RT chromosomal translocations in acute leukemias.";
RT Cell 71:691-700(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=96290553; Pubmed=8703835;
RA Nilsson I., Loechner K., Siegler G., Grell J., Beck J.D., Fey G.H.,
RA Marschalek R.;
RT "Exon/intron structure of the human ALL-1 (MLL) gene involved in
RT translocations to chromosomal region 11q23 and acute leukaemias.";
RT Br. J. Haematol. 93:966-972(1996).
RN [3]
RP SEQUENCE OF 1-1909 FROM N.A.
RX MEDLINE=93390935; Pubmed=8378076;
RA Yamamoto K., Seto M., Komatsu H., Iida S., Akao Y., Kojima S.,
RA Kodera Y., Nakazawa S., Ariyoshi Y., Takahashi T., Ueda R.;
RT "Two distinct portions of LTG19/BNL at 19p13 are involved in t(11;19)
RT leukemia.";
RT Oncogene 8:2617-2625(1993).
RN [4]
RP SEQUENCE OF 1317-2328 FROM N.A.
RC TISUE=Brain;
RX MEDLINE=93265134; Pubmed=1303259;
RA Djebail M., Salleri L., Parry P., Bower M., Young B.D., Evans G.A.;
RT "A trithorax-like gene is interrupted by chromosome 11q23
RT translocations in acute leukaemias.";
RT Nat. Genet. 2:113-118(1992).
RN [5]
RP SEQUENCE OF 1251-1538 FROM N.A.
RX MEDLINE=94215165; Pubmed=8162575;
RA Gu Y., Alder H., Nakamura T., Schlichman S.A., Prasad R., Canaanl O.,
RA Saito H., Croce C.M., Canaanl E.;
RT "Sequence analysis of the breakpoint cluster region in the ALL-1 gene
RT involved in acute leukemia.";
RT Cancer Res. 54:2326-2330(1994).
RN [6]
RP SEQUENCE OF 1251-1654 FROM N.A. (ISOFORM 14P-18B).
RX MEDLINE=95320205; Pubmed=7598802;
RA Mbongkolle D., Burnett R., McCabe N., Thirman M., Gill H., Yu H.,
RA Rowley J.D., Diaz M.O.;
RT "The human MLL gene: nucleotide sequence, homology to the Drosophila
RT trithorax-like domain, and alternative splicing.";
RT DNA Cell Biol. 14:475-483(1995).
RN [7]
RP SEQUENCE OF 1212-1603 FROM N.A.
RX MEDLINE=95315013; Pubmed=7794749;
RA Marschalek R., Grell J., Loechner K., Nilson I., Siegler G.,
RA Zweckbronner I., Beck J.D., Fey G.H.;
RT "Molecular analysis of the chromosomal breakpoint and fusion
RT transcripts in the acute lymphoblastic SEM cell line with chromosomal
RT translocation t(4;11).";
RT Br. J. Haematol. 90:308-320(1995).
RN [8]
RP SEQUENCE OF 1421-1540 FROM N.A.
RX MEDLINE=94020842; Pubmed=8414518;
RA Forster A., Rablitts T.H.;
RT "A method for identifying genes within yeast artificial chromosomes:
RT application to isolation of MLL fusion cDNAs from acute leukaemia
RT translocations.";
RT Oncogene 8:3157-3160(1993).
RN [9]
RP CHROMOSOMAL TRANSLOCATION WITH GAST7.
RX MEDLINE=20183971; Pubmed=10706619;
RA Megonigal M.D., Cheung N.-K.V., Rappaport E.F., Nowell P.C.,
RA Wilson R.B., Jones D.H., Addya K., Leonard D.G.B., Kushner B.H.,
RA Williams T.M., Lange B.J., Felix C.A.;
RT "Detection of leukemia-associated MLL-GAST7 translocation early during
RT chemotherapy with DNA topoisomerase II inhibitors.";
RT Proc. Natl. Acad. Sci. U.S.A. 97:2814-2819(2000).
RN [10]
RP CHROMOSOMAL TRANSLOCATION WITH AF3P21.
RX MEDLINE=20115194; Pubmed=10648423;
RA Sano K., Hayakawa A., Piao J.-H., Kosaka Y., Nakamura H.;

```

RT Novel SH3 protein encoded by the AF321 gene is fused to the mixed  
 RT lineage leukemia protein in a therapy-related leukemia with  
 RT t(3;11)(p21;q23).  
 RL Blood 95:1066-1068(2000).  
 CC -1- FUNCTION: POSSIBLY ACTS AS A TRANSCRIPTIONAL REGULATORY FACTOR.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- TISSUE SPECIFICITY: HEART, LUNG, BRAIN AND T AND B LYMPHOCYTES.  
 CC -1- DISEASE: INVOLVED IN ACUTE LEUKEMIAS BY CHROMOSOMAL TRANSLOCATIONS  
 CC T(11;19)(Q23;P13.3) THAT INVOLVES MLL AND MLT1/ENL;  
 CC T(4;11)(Q21;Q23) THAT INVOLVES MLL AND MLT2/AF4; T(9;11)(P22;Q23)  
 CC THAT INVOLVES MLL AND MLT2/AF9; T(6;11)(Q27;Q23) THAT INVOLVES  
 CC MLL AND MLT4/AF6; T(11;17)(Q23;Q21) THAT INVOLVES MLL AND  
 CC MLT6/AF17; T(X;11)(Q13;Q23) THAT INVOLVES MLL AND MLT7/AFX1;  
 CC T(10;11)(P12;Q23) THAT INVOLVES MLL AND MLT10/AF10;  
 CC T(1;11)(Q21;Q23) THAT INVOLVES MLL AND AF10; T(11;19)(Q23;P13.3)  
 CC THAT INVOLVES MLL AND ELL; T(11;19)(Q23;P23) THAT INVOLVES MLL  
 CC AND GAS7; and t(3;11)(p21;q23) that involves MLL and AF3p21.  
 CC -1- SIMILARITY: BELONGS TO THE TRANSCRIPTION FACTOR TRITHORAX FAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 BROMODOMAIN.  
 CC -1- SIMILARITY: CONTAINS 3 PHD-TYPE ZINC FINGERS.  
 CC -1- SIMILARITY: CONTAINS 1 CXXC-TYPE ZINC FINGER.  
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: L04284; AA58669.1; -;  
 DR EMBL: 269744; CA93625.1; -;  
 DR EMBL: 269745; CA93625.1; JOINED.  
 DR EMBL: 269746; CA93625.1; JOINED.  
 DR EMBL: 269747; CA93625.1; JOINED.  
 DR EMBL: 269748; CA93625.1; JOINED.  
 DR EMBL: 269749; CA93625.1; JOINED.  
 DR EMBL: 269750; CA93625.1; JOINED.  
 DR EMBL: 269751; CA93625.1; JOINED.  
 DR EMBL: 269752; CA93625.1; JOINED.  
 DR EMBL: 269753; CA93625.1; JOINED.  
 DR EMBL: 269754; CA93625.1; JOINED.  
 DR EMBL: 269755; CA93625.1; JOINED.  
 DR EMBL: 269756; CA93625.1; JOINED.  
 DR EMBL: 269757; CA93625.1; JOINED.  
 DR EMBL: 269758; CA93625.1; JOINED.  
 DR EMBL: 269759; CA93625.1; JOINED.  
 DR EMBL: 269760; CA93625.1; JOINED.  
 DR EMBL: 269761; CA93625.1; JOINED.  
 DR EMBL: 269762; CA93625.1; JOINED.  
 DR EMBL: 269763; CA93625.1; JOINED.  
 DR EMBL: 269764; CA93625.1; JOINED.  
 DR EMBL: 269765; CA93625.1; JOINED.  
 DR EMBL: 269766; CA93625.1; JOINED.  
 DR EMBL: 269767; CA93625.1; JOINED.  
 DR EMBL: 269768; CA93625.1; JOINED.  
 DR EMBL: 269769; CA93625.1; JOINED.  
 DR EMBL: 269770; CA93625.1; JOINED.  
 DR EMBL: 269771; CA93625.1; JOINED.  
 DR EMBL: 269772; CA93625.1; JOINED.  
 DR EMBL: 269773; CA93625.1; JOINED.  
 DR EMBL: 269774; CA93625.1; JOINED.  
 DR EMBL: 269775; CA93625.1; JOINED.  
 DR EMBL: 269776; CA93625.1; JOINED.  
 DR EMBL: 269777; CA93625.1; JOINED.  
 DR EMBL: 269778; CA93625.1; JOINED.  
 DR EMBL: 269779; CA93625.1; JOINED.  
 DR EMBL: 269780; CA93625.1; JOINED.  
 DR EMBL: D14540; BAA03407.1; -;  
 DR EMBL: L01986; AAA92511.1; -;  
 DR EMBL: U04737; AAA18644.1; -;  
 DR EMBL: S78570; AAB34770.1; -;  
 DR EMBL: X83604; CA58584.1; -;

DR EMBL: S66432; AAB28545.1; -;  
 DR EMBL: AF231998; AAG26332.2; ALT\_TERM.  
 DR TRANSFAC: T02337; -;  
 DR MIM: 159555; -;  
 DR InterPro: IPR001487; Bromodomain.  
 DR InterPro: IPR003889; Fyrich.C.  
 DR InterPro: IPR003888; Fyrich.N.  
 DR InterPro: IPR001965; PHD.  
 DR InterPro: IPR003616; PostSET.  
 DR InterPro: IPR001214; SET.  
 DR InterPro: IPR002857; znf-CXXC.  
 DR Pfam: PF00628; PHD; 3.  
 DR Pfam: PF00856; SET; 1.  
 DR Pfam: PF02008; zf-CXXC; 1.  
 DR SMART: SM00297; BROMO; 1.  
 DR SMART: SM00542; FYR1; 1.  
 DR SMART: SM00541; FYR2; 1.  
 DR SMART: SM00249; PHD; 4.  
 DR SMART: SM00508; PostSET; 1.  
 DR SMART: SM00317; SET; 1.  
 DR PROSITE: PS50014; BROMODOMAIN\_2; 1.  
 DR PROSITE: PS50280; SET; 1.  
 KW Proto-oncogene; Chromosomal translocation; DNA-binding; Bromodomain;  
 KW Nuclear protein; zinc-finger; Metal-binding; Transcription regulation;  
 KW Alternative splicing.  
 FT DOMAIN 17 102  
 FT DNA\_BIND 169 180  
 FT DNA\_BIND 217 227  
 FT DNA\_BIND 301 309  
 FT ZN\_FING 1147 1194  
 FT ZN\_FING 1431 1482  
 FT ZN\_FING 1484 1533  
 FT ZN\_FING 1566 1627  
 FT DOMAIN 1703 1748  
 FT DOMAIN 3840 3969  
 FT DOMAIN 137 143  
 FT DOMAIN 561 564  
 FT DOMAIN 568 571  
 FT SITE 1362 1363  
 Query Match 3.4%; Score 258.5; DB 1; Length 3969;  
 Best local Similarity 18.6%; Pred. No. 0.0005;  
 Matches 350; Conservative 243; Mismatches 611; Indels 673; Gaps 91;  
 QY 10 DPSRAETRRKRCPCDQGPSKRNTERKNEQENKYLEELAFANFDIDNFKRDK 69  
 DB 1673 NPETESIPRSRSPGDPPLVLTGVSKDDQPL-----DLEGVKKRMDQ 1717  
 QY 70 ---CALIK---ETVKQIQIRKQEKKAANID---EVQKSDVSSGCGVIDKDALGPMV 119  
 DB 1718 GNTSVLEFSDDIVKIIQ-----AINTSDGQPEIRKAN-----SM 1753  
 QY 120 LEALDFFEFVNVLEGNVVFSENVTOYLRYNOELNNKSVSYSLHVGDTPEYKN-LLPK 178  
 DB 1754 VKS-----FFIRQMERVFPPWVSKSRFMBENK-----VSSNGMLPNAVLPP 1796  
 QY 179 SI-VNGSGMSGEPPRRNSHTFNCRLVPLDSEEGHNDQEAHQYETMQCFVAVSQPS 237  
 DB 1797 SLDHNYAOW--QERENSHTEQPLMKKIIIPAKPKRGSGPDPRLPHPTPIILSDRS 1854  
 QY 238 IKREG-----EDLOSLICY-----ARVPKMERP-----VLPSSSEFTTR 273  
 DB 1855 REDSPELNPPGPIEDNRQCALCTYGDSDSANDAGRLLYIGQNEHTVNCALMSAEVFE 1911  
 QY 274 QDLOGKITSIDTSTMA-----AMKPGWE--DLVRRCIQKHQAHE-----GESV 316  
 DB 1912 EDDDGSLKNVHMAVIRKQLRCEFCOKPGATVGCCLTCTSNYHPCSRRAKNCVFLDDXK 1971  
 QY 317 SYAKRHH-----HEVLRQ-----GLAFSQIYRF----- 339  
 DB 1972 VYQGRHDLIKGEVPENGEVEFRVFDVEGISTLRKRLFLNGLEPPIHMMIGSMITDCL 2031  
 QY 340 ----SLSDGTLVAQTK-----SKLIRSQTTEPQVLVLSLMLHREQNVCVNMPDLT 387



```

Db 2032 GILNDLSD-----CEDKLEPIGYOCSRYWSTTDARKRCVYCKIV--ECRPVVEPDI- 2083
Oy 388 GQTMKPLNPFISSNPAHALCSCNPGQDMITSSNINPFIINPKRQMGPMKRFQSSGM 447
Db 2084 NSTVEHDEKRTIAHSPSTESSKSESONTA-----EIISSPDPDRPHSGTSCCY 2136
Oy 448 NHVSGM-QATTPQSGNYALKMNSPQSSPGMNP-----GQPTSLMSLRHR-----MSR 494
Db 2137 HVIKSVPRIRTPS-----YSPQRSGCRPLPSAGSPT---PTHETIYVGPDLSS 2185
Oy 495 GY--AGSPRIIPSPSPACS---LHSPVGCSTGNSHSYTNS-----533
Db 2186 GLRSIGSRHSTSLSPQSKLRIMSPM---RTGNTYSRNNVSVSTGTATDLESSAK 2241
Oy 534 -----LINAQALSEGCVGLSSLASPDLMKMLQNSPNNMPPRLSKNGS---LDS 582
Db 2242 VVDHVLGPNLNSSTSL--GGNTSTSSMLQRTVTVGN--KNSHLDGSSSSMKOSSASDLVS 2298
Oy 583 KDCFGIYGE-----PSEGTGQAESCHPEQEKETNDPMLPPAVSSERADGOSRLHDSKG 637
Db 2299 KSS--SLKGEKTKVLSSKSEGAHNAVAYPG-----IPLKAP-----2333
Oy 638 QTKLIQLTTKSDQMEPPLA-----SSLSDTNKOSTGSLPGSGSTGTSLKEKHILHR 692
Db 2334 -----QVHNTTSRELNVSIGSFAPSSVSFSSKE-----ALSFPHLHLRG 2374
Oy 693 LLQDSSPVDLAKLAETAKGDLSSSSTAGSEVTIKQEVSPKKKNALLRYLLMD 752
Db 2375 QNRNDQHTD-----STQSANSPDEDEVTYTKLGSNNSSTI-----2414
Oy 753 DTKDIGLEPITKLERLDSKTPASNTKLIAMKTEKEEMS---PEPQDQSELDNLE- 807
Db 2415 -NEHMG-----SSRRROKRGKSKCKETREKHSKSFLEPGVTTGEBENLKP 2462
Oy 808 EILDLQNSQLPOLFPD-----TRGPAP-----AGSYD-----835
Db 2463 EEMDEV-----LTPYWGQRPCCNWNSSDKIGDKGLMGPVKAPPMQVEGSAKELQAP 2515
Oy 836 -KQALINDLMQLTAE-----SPVTVGAKQKTLRLISQSTFNNPRGQGLRLLP 883
Db 2516 KRRYKVTTLTPLKMNESOKNALESSPASPLQIESTSPTEPISASENPDGPAQPS 2575
Oy 884 N-----QNLPL-DITLQSPGTAGP-----FP-----PIR 906
Db 2576 NNTSCQDOSNNYQNLPLPQDRMLMLPDGPKPQEDGSFKRRYRRSARASNMFFGLTPLY 2635
Oy 907 NSSPY-----SVLPQGMGNQGMIGNQNLGNS-----STGMIG 941
Db 2636 GVRSYGEEDIPYSSSTGKRGKRSAGQVDGADLSTDEDDLYYNNFTIVISSGGE 2695
Oy 942 NSASR-----PRMPSGEMAPQSSAVRYTCATTSAMNR--PVQGMIRNPASTP 989
Db 2696 RLASHNLFREBQCDLPKISQLDGVDGTESDTSVTATTRKSSQILPKRNGK-ENGTEMLK 2754
Oy 990 MRPSQPOROTLO-----SOVANNIGPSELEMMGPGPOYSO-----1025
Db 2755 IDRPDAEKEHVTSSVGHKNEPKMKNCHSVKTKQGOQDLEQLSLESSRKYHST 2814
Oy 1026 -----QOAPNPOTAPWPEIILPIDQASFASQN---ROPGSSPDLLCPH 1067
Db 2815 PSDKMLDPTYNTELKSSDNNNSDDCGNILLPSDIMDFLKNTPPSMAGLGEPS-----2867
Oy 1068 PAESPSPDEGALLDQLYALRNFDGLF-EIDRALGIPELVSSQAVDPQFSSQDSNIML 1126
Db 2868 ---ESSSSB-----LNLGEGLDISNREKDMGLFEVFSQQLPTTEPVDSSVSSISA 2917
Oy 1127 BO-----KAPVPOQYASQOMAGSVSPQODNFHMGOR-----P 1163
Db 2918 EQQFELPLEPLPSDLVLTTRSPVPSQNPRLAVISDS-----GKRYVTIEK 2965
Oy 1164 SVATLRMP-----RPGLRPTGLVQNOPNQLRLQRLQAOQNRPLMNOI--SNVSNVL 1218

```

```

Db 2966 SVASSESDPALLSPGVDPTRPEGHMPDHF--IQGHMADHISPPCGSVQGHGNDDL 3022
Oy 1219 TLRPGVP--TQADIPNQMOLAQOREILNQ-----HLR---OROMHQ 1255
Db 3023 TRNNSSTPGLOVEVSPVPIQONOKIVPNSDTSQPSQISANAQVTPPHILKPAETKLIYVN 3082
Oy 1256 QOVQOQ-----RTL-----MARGOGLNMT-----SMVASGPMATMS---NPRIPA 1294
Db 3083 QMMQPLVYLQTLPNQYQKIQTLSSVSSPVSVMENRTSVLGMGGGLTLITGLNPLSPIS 3142
Oy 1295 N-----AQOPFPN-----YGISQOPDGF-----1315
Db 3143 QSLFSPASKGLLPMSHHQHLHSPPAATOSSFPFNISNPPGILIGVQPPDQLLYSESS 3202
Oy 1316 -----TGATTPQSPMLSPRMAHTQSPMMQO---SQANPAYQAPSDI-----1353
Db 3203 QRTDSTVATPVSGLKRPISRLQTRKKKLAPSPSTPINAPSDVSNMTLINFTPSOL 3262
Oy 1354 -NGMAQNGMNGSMFESQSPHFQOQANTS-MYSNNMNIIVSMATNTGQSSMNOITQOI 1411
Db 3263 PNHPLSLDLGSLNTSSHRTVPNIIKRSKSIWPEPAPL---LPQSVGCTAFAAGTSTI 3319
Oy 1412 SM--TSVTSVSTGLSS 1426
Db 3320 SODTSHLTSGSVGLAS 3336

RESULT 28
MOZ_HUMAN STANDARD; PRT; 2004 AA.
ID 092794;
AC 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DI 16-OCT-2001 (Rel. 40, Last annotation update)
DE Monocytic leukemia zinc finger protein (zinc finger protein 220).
GN ZNF220 OR MOZ.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN 11)
RP SEQUENCE FROM N.A.
RX MEDLINE=96376968; PubMed=8782817;
RA Borrow J., Stanton V.P., Jr., Andresen J.M., Becher R., Behn F.G.,
RA Chaganti R.S.K., Civin C.I., Distche C., Dube I., Fritschauf A.M.,
RA Hossain D., Mittleman F., Volinia S., Watson A.E., Housman D.E.,
RT "The translocation t(8;16)(p11;p13) of acute myeloid leukemia fuses
RT a putative acetyltransferase to the CREB-binding protein.";
RL Nat. Genet. 14:33-41(1996).
CC -!- FUNCTION: MAY REPRESENT A CHROMATIN-ASSOCIATED ACETYLTRANSFERASE.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- DISEASE: PARTICIPATES IN A T(8;16)(P11;P13) CHROMOSOMAL
CC TRANSLOCATION THAT PRODUCES A MOZ-CBP CHIMERA OBSERVED IN THE
CC M4/M5 SUBTYPE OF ACUTE MYELOID LEUKEMIA.
CC -!- SIMILARITY: CONTAINS 2 PHD-TYPE ZINC FINGERS.
CC -!- SIMILARITY: BELONGS TO THE MYST (SAS/MOZ) FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U47742; AAC50662.1; -.
CC HSSP: 060631; 4GBQ.
CC MIM: 601408; -.
CC InterPro: IPR001386; Linker_histone.
CC InterPro: IPR002717; MOZ_SAS.
CC InterPro: IPR001965; PHD.
CC Pfam: PF01853; MOZ_SAS; 1.
CC Pfam: PF00628; PHD; 2.

```



```

CC -1 SUBCELLULAR LOCATION: Nuclear.
CC -1 SIMILARITY: CONTAINS 1 BROMODOMAIN.
CC -1 SIMILARITY: CONTAINS 1 ZN-TYPE ZINC FINGER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: S66385; AAB28651.1; -
DR TRANSFAC: T01318; -
DR MGD: MGI:1098280; Crebpb.
DR InterPro: IPR001487; Bromodomain.
DR InterPro: IPR001301; KIX.
DR InterPro: IPR00197; TAZ_finger.
DR InterPro: IPR000433; ZnF_ZZ.
DR Pfam: PF00439; Bromodomain; 1.
DR Pfam: PF02172; KIX; 1.
DR Pfam: PF02135; ZF-TAZ; 2.
DR Pfam: PF00569; ZZ; 1.
DR PRINTS: PR00503; BROMODOMAIN.
DR SMART: SM00297; BROMO; 1.
DR SMART: SM00291; ZnF_ZZ; 1.
DR PROSITE: PS00633; BROMODOMAIN_1; 1.
DR PROSITE: PS50014; BROMODOMAIN_2; 1.
DR PROSITE: PS01357; ZF_ZZ_1; 1.
DR PROSITE: PS50135; ZF_ZZ_2; 1.
DR Transcription regulation; Nuclear protein; Activator; Bromodomain;
KW Zinc-finger.
FT DOMAIN 1104 1176 BROMODOMAIN.
FT ZN_FING 1702 1745 ZZ-TYPE.
FT DOMAIN 1062 1065 POLY-GLU.
FT DOMAIN 1556 1563 POLY-GLU.
FT DOMAIN 1944 1949 POLY-PRO.
FT DOMAIN 1968 1971 POLY-GLN.
FT DOMAIN 2082 2086 POLY-GLN.
FT DOMAIN 2200 2216 POLY-GLN.
FT DOMAIN 2296 2299 POLY-GLN.
FT SEQUENCE 2441 AA; 265474 MW; 0AB8028C3112F419 CRC64;
SO QUERY MATCH
Best Local Similarity 3.3%; Score 252; DB 1; Length 2441;
Matches 248; Conservative 142; Mismatches 442; Indels 372; Gaps 61;
QY 392 GKPLNP-----ISSNSPAHQALCSGNPGQDMTLSSNINFPINGPKREMGMPKRGFGSGGM 447
DB 78 GSSINRPGIGNVASASSPVQGG- -GGQAQGGQPNSTN-----MASLGAM-GK 119
QY 448 NHVSOMQATTPOGSNTALKMNSPSSQSPGMNPGOPTSM-LSFRHRKNSPCVACSPRIAPSQ 506
DB 120 SPLNGDSSSTP-----NLPKQAASSTGPTPPASQALNPOAQKQVGLVTS----- 163
QY 507 FSPAGLSLSPVVCSSSTGNSHSY-----TNSSLNALQALSEGHSVLSGLSLASPLDKMN 561
DB 164 -SPATISQTPG- GICMNAFNPQTHPGLINSNSGSHLMOAQOQAOVYNGSL----- 212
QY 562 LQNSFVNMPPLSKMGLSKDCFLYGEPSGTTGQAESSCHPGEOKETNDPNLPAPV 621
DB 213 -----GAAGRGRCGAGMPYPA-----PAMQCAT 234
QY 622 SSERAD-----GOSRLHDSK--GOTKLLQLITTSDDMEESPPLASSISLDRKSDST 669
DB 235 SSVALETLTVQVSPQMAHAGLNTAQAGGTTKMGMTGTT-----SPFGQPSQTGGQOM 287
QY 670 GSLPSSGTHGTHSLKEKHILHRLLODSSSPVDLAKLTAEATGKDLSDQSSSTAPGSEVT 729
DB 288 GA-----TGVNPOLASKOSMNSL---PAFPDI- KNTSVTTVPNNKSLQTSVGLVPTQA 338
QY 730 IKQEPVSPKKENALLRYLLDKDTRKIDGLPEITPKLERLDSKTDPSANTKLIAMKTEKE 789

```

```

DB 339 IATGPFADEKERRLLQOQLV-----LLJHAHQCRR 369
QY 790 EHSFERGDDPGSELND---LEELIDLONSQLPQLFPDTPRGPAPAGSVKQALINDMLQ 846
DB 370 EDA--NGEVRACSLPRCPTKKNVLYNHTHCQAP-----KACQVHACASSROTIHMKNC 421
QY 847 TAEHSVFPVFGAOKTALRISOSTFNNRPGQLGRLLPNQNLPLDTLTLQSPGAGPPPIR 906
DB 422 TTHDCV-----CLPKNADSKRNRQITLG----- 446
QY 907 NSSPYSLVTPQGMNGOIGNOGNGSSTGMINGSARLPMPSGEMAPQSSAVRVTCA 966
DB 447 --SPAS-----GIQNTIGSVGAGQGNATSL--SNBNPTDP-----SSMGAYAA 486
QY 967 ATTSAMNRPVQGMIRNPAASIIMRPSQPGQR- -TQSOVYN-IGPSLEEMMGPOYS 1024
DB 487 LGLPYKNOP-----OTQLOPYVPGQOPAPQAPHAQKRTLALGNNMVSYPAGITTD 538
QY 1025 QOQAPNPQTPAPWPESTILPIDQASFASQNRQP- -GSSPDDL-LCPHRAESSPDEGA- 1078
DB 539 QQ--PPNLLS-----ESALP---TSLGATN--PLMNGSNSNGNIGSLSTIPTAAPSSGTGR 588
QY 1079 -----LLDQLYLAL-RNFDGLEETIDRALGIPELVQSQAADPDQFSSQDSN 1123
DB 589 KGMHEHVTDRLSRHLVHKLVQAIFPTPDPAALKDRM--ENLVAAKVEGDMYESANSR 646
QY 1124 ----IMLEQKAVFVPOQVYSQAQMAQGSYPPQDPNFTMGQRPSTATLRMPRRLART 1179
DB 647 DEYHLLAKKI-----YKIOKELEEKRRTRLRKQGI--LGNOPALPASQAP-PVIRPA 697
QY 1180 GLVONOPQRLRLQLOHRLQAOONRQPLMNQISNVSNVNTLR----- 1221
DB 698 QSVRRPNGLPLPV--NRMOVSOG-----MNSFPMISGNVQLQAPRGPRAASPMHNSVOM 752
QY 1222 -----PGV--PTQAPINAOMLAQOREILNOROMHOQOQVQFTLMR-----GQ 1268
DB 753 NSMASVPGMATSPSRMPPQPNMNGTHANNIMAOAPLQNOFLPQNOFPSSGAMSVSVGM 812
QY 1269 GLNMTSPMAVPSGMP-ATMSNP---RIPOANQOQFPFPNYTGISOQ----- 1311
DB 813 GPPAAQAGVSQOGEPPAALPNPLMLAPQAS--QLPCRP---VTOSPLRTPPPASTAAG 867
QY 1312 -----DPGFT-----GATTPOSPLMS-----PRMAYTOS-PMQO--QSOA 1343
DB 868 MSLQHPHTAPKMTPPQPAAPTPQSTPVSQGTPTTPGVSYPAAQTQSTPTVQAAQAQV 927
QY 1344 NPAYQAPSDINDMAQGNMGNSMFSQSP--PHFGQANTSMYNNNNIN---VSNATNTG 1399
DB 928 TPQQPTP-----VQPPSVATPQSSQOQPTPVHTQPGTRPLSQAAASIDNRVPTSTVRS 981
QY 1400 GMSNMQNTGQJSMYSV--TSVSTSLGSSMGPRQVNDPALRGNLPLPNQLRGMDMKQED 1458
DB 982 AETSSQPPPDVPMLEMKTEVQTDAPPTESKBP--SEMMEDLOGSSOVAKETD 1038
QY 1459 TTRK 1462
DB 1039 TTEQ 1042

```

```

RESULT 30
ID YAV1_SCHPO STANDARD: PRT: 1794 AA.
AC 010172:
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Hypothetical 193.3 kDa protein C27F1.01c in chromosome I.
GN SPAC27F1.01C OR SPAC25G10.09C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.

```



```

CC CC -1- FUNCTION: MAY HAVE A REGULATORY FUNCTION POSSIBLY IN ASSOCIATION
CC CC WITH THE N GENE PRODUCT.
CC CC -1- SUBCELLULAR LOCATION: Nuclear.
CC CC -1- DEVELOPMENTAL STAGE: DURING EARLY NEUROGENESIS MAM PRODUCTS ARE
CC CC UBQUITOUSLY LOCATED. DURING LATER STAGES THEY ACCUMULATE IN THE
CC CC CENTRAL NERVOUS SYSTEM.
CC CC -1- MISCELLANEOUS: THE PROTEIN HAS MANY AA HOMOPOLYMERIC DOMAINS: 21
CC CC POLY-GLN RUNS (FROM 5 TO 16 AA IN LENGTH), 4 POLY-GLY (6 TO 10
CC CC AA), 3 POLY-ASN (3 X 5 AA), 1 POLY-ALA (10 AA) AND 1 POLY-THR (5
CC CC AA) RUNS.
CC CC -1- SIMILARITY: TO OTHER NUCLEAR PROTEINS OF DROSOPHILA, TO CERTAIN
CC CC YEAST AND MAMMALIAN REGULATORY PROTEINS.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL, X54251, CA938152.1; -.
CC CC PIR, A33106; A33106.
CC CC PIR, A36391; A36391.
CC CC FlyBase: FBgn0002643; mam.
CC CC DR Neurogenesis; Nuclear protein. Repeat.
CC CC KW DOMAIN 20 84 GLN-RICH.
CC CC FT DOMAIN 127 190 ARG/LYS-RICH (BASIC).
CC CC FT DOMAIN 196 219 GLN-RICH.
CC CC FT DOMAIN 259 304 ASN-RICH.
CC CC FT DOMAIN 355 388 GLY/ASN-RICH.
CC CC FT DOMAIN 392 406 GLN-RICH.
CC CC FT DOMAIN 407 440 GLY-RICH.
CC CC FT DOMAIN 651 671 GLN-RICH.
CC CC FT DOMAIN 700 714 GLN-RICH.
CC CC FT DOMAIN 759 816 GLN-RICH.
CC CC FT DOMAIN 987 996 5 X 2 AA TANDEM REPEATS OF G-V.
CC CC FT DOMAIN 1060 1079 ALA-RICH.
CC CC FT DOMAIN 1092 1107 8 X 2 AA TANDEM REPEATS OF V-G.
CC CC FT DOMAIN 1237 1252 7 X 2 AA TANDEM REPEATS OF G-V.
CC CC FT DOMAIN 1492 1496 POLY-THR.
CC CC FT DOMAIN 1559 1592 ASP/GLU-RICH (ACIDIC).
CC CC SQ SEQUENCE 1596 AA; 167717 MW; B944D86EF359D605 CRC64;
CC CC -----
CC CC Query Match 3.3%; Score 250; DB 1: Length 1596;
CC CC Best Local Similarity 19.8%; Pred. No. 0.00038;
CC CC Matches 228; Conservative 112; Mismatches 449; Indels 360; Gaps 49;
CC CC -----
QY 472 QSSPQNPQGTSMLSPPHRRMSPVAGSPRIIPPSQFSPAGSLHSPVGCSSSTGNSHYTN 531
DB 216 QQAQTMLAGQGLSSVHYGQKFLKRAEDVDNGPDSFEFRPHKL-----PNNNNNSN 267
QY 532 SLLNALQALSGHGVSGLSSLASPDLMKGNLONSFPVNNAPRLSKMGSLDKDFGLYGE 591
DB 268 NNNGNANANNNGGNSNTNNT-----NNNGSTNNNG-----GS 301
QY 592 PSEGTGAESSCHRGDKE--TNDPNLPAVSSSRADQGSRLHDSKGTATLDTTKS 649
DB 302 NNNGSENITKFSVELVQDLEFTTSPANSOPQOISTNV-----TVKALTNTS 347
QY 650 DQMEP-----SPLASSLSDTNKSTGSLSPGSGTGT-----SLKERKHLIARLQDS 697
DB 348 VKSEPGVGGGGGGGGGSGNNNNNNNGGGGGGNNNNNGGDNHQQDQHQDQQDQ 407
QY 698 SSPVDLALITLFAHGKDLDSQSSSTAPRSEVTIKQEPVSPKKKNAALLRYLDKDTKDI 757
DB 408 GGLGLGN-----NRRGGGPGGMAIGPGVA-----GGLGGM 439
QY 758 GLPFTPLERLDSKTDPASTKLIAMTEKE-----EMSFEPGDPQSGELMLEITLDL 813
DB 440 GMP-----PMMMSAQKQASALGNLANILVECKRREPDRHDPDLGSLDKDGGGGGPFEPDLGD- 495

```

[illegible]



QY 260 EREVLDSSEFTTRDLOG-KITSLDTST-----MRAAMKPGMEDLVRCLOCKFHAONE 312  
 Db 2488 PRPOCPPLQSSPSPQLSHLPLKPLHSTRPQOLANLPOLIPVQCGCKLAPSPFHEMGE 2547  
 QY 313 GSEVSTAKHHHEVLRQGLAFSOTYRFSLSD--GTLVAQOTKSKLIRSOTTEPOLYISL 370  
 Db 2548 HOOLHLSAONOFIHPOLFDRSLDMFMLEPSPNPLASOLLSCAI-----FOIPAS- 2599  
 QY 371 HMLHRONCVMPNPDLTGOTMGKPLNPISNSPAHQALCSG--NPQODMTLSSINPFIN 428  
 Db 2600 -----SATSPSTPTMTMTLTKRLLEKASASGENDSGTGEEPODKRLRTTIT---- 2649  
 QY 429 GREQOMGMBGRFGSGG-----MNHVS---GMOATTPQGSNYALKMNSPQSSPG---- 476  
 Db 2650 --PEQELIYOKYKLDNSNTRKMLDIAHEVGLKRRVYO---VWFQNTFRARRKQOFRA 2703  
 QY 477 MNPQOPTSMLSPPHMSR-----GVAGSPRIPPSQFSPA-----G 511  
 Db 2704 VGPAAQ-----AHRRCPCRCALFKAKTALAEHIRSRHHEAKRAGYNLTLSMLLDCDG 2756  
 QY 512 SLHSPVYGVCSSTGNSHTYNSLSNALQALSEHGVSIG-----SLASP-DLXMG 560  
 Db 2757 GLQMGKIDTGTFSHLPSS-----SDQGVPLSPVSKTMEIPLSLSPSSITKVE 2808  
 QY 561 NLQN-----SPVNMNPPPLSKMGLDKCEGLYGEPESEGT---GOAESCHPG---E 608  
 Db 2809 GJEDFESPMSSVNLN-----FDQTKLDNDCCSVNTAITDTTGDGSDMNDNATGATE 2864  
 QY 609 QETNDPN--LPPAVSSEADGOSRL-----HDSKOTKILLQLLTKSDQ 651  
 Db 2865 TKSSAPNNGELTKAAMAMASEYEDRLSGLVSPAPSFYSKEYNEGTVDYSE--TSSLADP 2923  
 QY 652 MEPSPLASSLDPNNDSTSLPGSGSTHGTSLKEKHKILHRLLODSSPDLAKLAENAT 711  
 Db 2924 CESPSPASSAGKSGDS--GDRPOQKFRFQOMTNLQKLVKSCNDNRTP---TMLCEVL 2979  
 QY 712 GRDLSESSSTAPGSEVTKOEPVSPKKENALRYLRLDKDTRKDIGLEPIT---FRL- 767  
 Db 2980 GMDIG-----LPKRVYQVWFQNAKRAKESKRL-----SMAKHNGINTSVEGKTEC 3026  
 QY 768 -----RLDSKTPASNTKLIAMKTEKESFEFGQPGSELNLEIIDLONSOL 818  
 Db 3027 TLGKIYSARLSVRDHIFSQOHISKVK-----TLTGSQDLKEKEFEPATVRQL 3075  
 QY 819 POLFPTRGAPAGSVDAKQAITNDLMQTLAENSPV---TPVGAOKTALRISOSTFNPR 875  
 Db 3076 -----MAQELDRITKKAHEVLGLAAQOQGMFNDTPLDA-----LMLPTA 3114  
 QY 876 GOLGRLLPNQNLPLDTTLOSPTGAGFPPIRNSPYSVPIPOPGMGNOGMIGNGLNS 935  
 Db 3115 YFALQGIPLVLP--GLNSPLPG-FTP--SNTALTSKPLM----- 3152  
 QY 936 STGMIGNSASRPPTMPSGEMAPOSSAIVTCATTSAMNRPVQGMIR-NPA-ASIPWRBS 993  
 Db 3153 --GLSTYTPSPGLP-----TSGLPNKPSSASLSSPTPAQITMAMGP- 3192  
 QY 994 SOPGQROTLOSQVMNIGPSELEMMNGRQYSSQQAARPNOTAPRESILPIDQ----- 1045  
 Db 3193 QQPPOQOQOQOQO-----POV-QQPPPPPAQPPPTPOLPQOQOQORRDXD 3236  
 QY 1046 ASFASONROPFSSPDDLCP--HPAESPDEGALLDOLYLARNFDGLEIDRALGIFE 1104  
 Db 3237 SEKVEKEKAHKGKGRPLLVPRKKEGEAPTATATIS-----AP 3275  
 QY 1105 LYSQSOAVDPEOPS-----SODSNIMLEOK-----APVFPQOYASQAOA-OGSISPMQ 1152  
 Db 3276 LPTMEYAVDPAQLOALQALTSPTALILTSQFLPYFVPGFSFYAPQIFGALQSGY---L 3332  
 QY 1153 DNEFHTMGORPSVATLRMPRGRLRPTGLVONQPNQLRLQLHRLQAQONRQPLMJOIS 1212  
 Db 3333 QPMTGEGELFP-YSPALSOALMGLSGSLDQ-----OYOOTQOSL----- 3371  
 QY 1213 VSNVNLTRPGVPTQAPINAOMLAORREILNDLROBOMHQOQOYQOORTIMARGOGLMN 1272

Db 3372 -----QEA10QOQOQORLQOQOQOQOQOQO--QPKASQTPV 3403  
 QY 1273 TFSMVAFSGMPATWSNPRIPQANAQOFPPPPNNGISQOQDPDGTGATTPQSPLMASRMAH 1332  
 Db 3404 PRGAPSPDKDPAKESPKREBKNTPREVSPPLPKLPEEPKESKADSLDPIVPRVOY 3463  
 RESULT 33  
 CLOC\_DROME  
 ID CLOC\_DROME STANDARD: PRT: 1023 AA.  
 AC 061735; 076342; 077137;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Circadian locomotor output cycles kaput protein (dClock) (dPAS1).  
 GN CLK OR JRK OR CLOCK OR PAS1.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX MBL\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-Head.  
 RX MEDLINE=98279147; PubMed=9616122;  
 RA Darlington T.K., Wager-Smith K., Ceriani M.F., Staknis D., Gekakis N.,  
 RA Steeves T.D.L., Weitz C.J., Takahashi J.S., Kay S.A.;  
 RT "Closing the circadian loop: CLOCK-induced transcription of its own  
 RT inhibitors per and tim.";  
 RL Science 280:1599-1603(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A., AND MUTAGENESIS.  
 RC TISSUE-Head.  
 RX MEDLINE=98292177; PubMed=9630223;  
 RA Allada R., White N.E., So W.V., Hall J.C., Rosbash M.;  
 RT "A mutant Drosophila homolog of mammalian Clock disrupts circadian  
 RT rhythms and transcription of period and timeless.";  
 RL Cell 93:791-804(1998).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CANTON-S;  
 RX MEDLINE=98414630; PubMed=9742131;  
 RA Bae K., Lee C., Sidote D., Chuang K.-Y., Ederly I.;  
 RT "Circadian regulation of a Drosophila homolog of the mammalian clock  
 RT gene: PER and TIM function as positive regulators.";  
 RL Mol. Cell. Biol. 18:6142-6151(1998).  
 CC -1- FUNCTION: CIRCADIAN REGULATOR THAT ACTS AS A TRANSCRIPTION FACTOR  
 CC AND GENERATES A RHYTHMIC OUTPUT WITH A PERIOD OF ABOUT 24 HOURS.  
 CC OSCILLATES IN ANTIPHASE TO THE CYCLING OBSERVED FOR PERIOD (PER)  
 CC AND TIMELESS (TIM). ACCORDING TO REF.3, REACHES PEAK ABUNDANCE  
 CC WITHIN SEVERAL HOURS OF THE DARK-LIGHT TRANSITION AT ZTO  
 CC (ZEITGEBER 0), WHEREAS REF.1 DESCRIBES BIMODAL OSCILLATING  
 CC EXPRESSION WITH MAXIMUM AT ZT5 AND ZT23. CLOCK-CYCLE HETERODIMERS  
 CC ACTIVATE CYCLING TRANSCRIPTION OF PER AND TIM BY BINDING TO THE E-  
 CC BOX (3'-CAAGCG-5'), PRESENT IN THEIR PROMOTERS. ONCE INDUCED,  
 CC PERIOD AND TIMELESS BLOCK CLOCK'S ABILITY TO TRANSDUCATE THEIR  
 CC PROMOTERS.  
 CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER  
 CC BHLH PROTEIN. FORMS A HETERODIMER WITH CYCLE.  
 CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).  
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: THE FULL-LENGTH VARIANT A  
 CC (SHOWN HERE) AND VARIANT B. ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC VARIANT B ENCODES TWO CONCEPTUAL PROTEINS, THE FIRST CONSISTS ONLY  
 CC OF THE BHLH DOMAIN, THE OTHER CONSISTS OF THE PAS-1 AND ALL C-  
 CC TERMINAL DOMAINS. VARIANT B IS EXPRESSED WEAKLY AT ALL THE TIME OF  
 CC THE DAY, AND IT CYCLES IN PHASE WITH THE FULL-LENGTH FORM.  
 CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED. FOUND IN HEAD, BODY, AND  
 CC APPENDAGE FRACTIONS.  
 CC -1- DOMAIN: CONTAINS THREE POLYGLUTAMINE REPEATS WHICH COULD  
 CC CORRESPOND TO THE TRANSDUCATION DOMAIN. THE LENGTH OF THE  
 CC REPEATS IS POLYMORPHIC. IN THE ARRYTHMIC MUTANT JRK, DELETION OF  
 CC THIS REGION LEADS TO THE LOSS OF CIRCADIAN RHYTHMICITY AND ALTERED







QY 821 L--FP 823  
DB 782 LTMFP 786

RESULT 35  
HMCU\_DROME  
ID HMCU\_DROME STANDARD; PRT; 2175 AA.  
AC P10180: 09W306;  
DT 01-MAR-1989 (Rel. 10, Created)  
DT 01-MAR-1989 (Rel. 10, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Homeobox protein cut.  
GN CT OR CG11387.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_Taxid=7227;  
RN SEQUENCE FROM N.A.  
RX MEDLINE=88232956; PubMed=2897632;  
RA Blochlinger K., Bodmer R., Jack J., Jan L.Y., Jan Y.N.;  
RT Primary structure and expression of a product from cut, a locus  
RL involved in specifying sensory organ identity in Drosophila.";  
RL Nature 333:629-635(1988).  
RN [2]  
RP SPOURCE FROM N.A.  
RC STRAIN=BERKELEY;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers J.H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Baeson K.Y., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,  
RA Burks K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Foslcr C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Malsbina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclby J.M.,  
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,  
RA Williams S.M., Woodage T., Worley R.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster.";  
RL Science 287:2185-2195(2000).  
CC -1- FUNCTION: REGULATOR OF CELL FATE DECISIONS IN MULTIPLE LINEAGES.  
CC SPECIFICALLY, FUNCTIONS AS A DETERMINATION FACTOR THAT SPECIFIES  
CC SENSORY ORGAN IDENTITY IN PRECURSOR CELLS. PROBABLY ALSO INVOLVED  
CC IN CELL TYPE SPECIFICATION OF MALPIGHIAN TUBULES. IN ABSENCE OF

CC CUT GENE EXTERNAL SENSORY ORGANS ARE TRANSFORMED INTO CHORDONAL  
CC ORGANS.  
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).  
CC -1- TISSUE SPECIFICITY: DETECTED IN MANY CELLS IN THE CENTRAL NERVOUS  
CC SYSTEM, ALL EXTERNAL SENSORY ORGANS, SOME PERIPHERAL NEURONS, AND  
CC IN THE NON-NEURAL CELLS OF THE SPIRACLES AND THE MALPIGHIAN  
CC TUBULES.  
CC -1- DEVELOPMENTAL STAGE: CELL-SPECIFIC PATTERN OF EXPRESSION. BROADLY  
CC EXPRESSED DURING EMBRYONIC DEVELOPMENT.  
CC -1- DOMAIN: ASN AT POSITION 47 OF THE HOMEBOX MAY PARTICIPATE IN  
CC REGULATING DNA-BINDING ACTIVITY BY PROMOTING HOMO- AND  
CC HETERO-DIMERIZATION.  
CC -1- SIMILARITY: BELONGS TO THE CUT FAMILY OF HOMEBOX PROTEINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: X07985; CA30794.1; -.  
DR EMBL: AE003441; AAF46264.2; -.  
DR PIR: S03170; S03170.  
DR HSSP: P04002; IWFA.  
DR TRANSFAC: T02004; -.  
DR FlyBase: FBgn0004198; ct.  
DR InterPro: IPR003350; CUT.  
DR InterPro: IPR000047; HTH\_repressr.  
DR InterPro: IPR001356; Homeobox.  
DR Pfam: PF002376; CUT; 3.  
DR Pfam: PF00046; homeobox; 1.  
DR PRINTS: PR00031; HTHREPRESSR.  
DR SMART: SM00389; HOX; 1.  
DR PROSITE: PS00027; HOMEBOX\_1; 1.  
DR PROSITE: PS50071; HOMEBOX\_2; 1.  
KW Transcription regulation; Homeobox; DNA-binding;  
KW Developmental protein; Nuclear protein; Repeat; Coiled coil.  
FT DOMAIN 265 343  
FT COILED COIL (POTENTIAL).  
FT DOMAIN 433 499  
FT COILED COIL (POTENTIAL).  
FT DNA\_BIND 798 964  
FT COILED COIL (POTENTIAL).  
FT DOMAIN 1056 1161  
FT COILED COIL (POTENTIAL).  
FT DNA\_BIND 1329 1417  
FT COILED COIL (POTENTIAL).  
FT DOMAIN 1463 1522  
FT COILED COIL (POTENTIAL).  
FT DNA\_BIND 1608 1695  
FT COILED COIL (POTENTIAL).  
FT DOMAIN 1745 1804  
FT COILED COIL (POTENTIAL).  
FT DOMAIN 194 210  
FT COILED COIL (POTENTIAL).  
FT DOMAIN 235 293  
FT COILED COIL (POTENTIAL).  
FT DOMAIN 271 293  
FT COILED COIL (POTENTIAL).  
FT DOMAIN 347 428  
FT COILED COIL (POTENTIAL).  
FT DOMAIN 547 584  
FT COILED COIL (POTENTIAL).  
FT DOMAIN 616 630  
FT COILED COIL (POTENTIAL).  
FT DOMAIN 665 699  
FT COILED COIL (POTENTIAL).  
FT DOMAIN 2004 2014  
FT COILED COIL (POTENTIAL).  
FT DOMAIN 2071 2077  
FT COILED COIL (POTENTIAL).  
FT DOMAIN 2124 2136  
FT COILED COIL (POTENTIAL).  
SQ SEQUENCE 2175 AA; 233628 MW; 08BF00C4861BD0AB CRC64;

Query Match 3.2%; Score 248; DB 1; Length 2175;  
Best Local Similarity 19.2%; Pred. No. 0.00069;  
Matches 305; Conservative 202; Mismatches 669; Indels 416; Gaps 68;

QY 7 NTSDPSRAETRRKRECPDQIGPSFKRTEKRNDEQKTYIEELAEILFANFNIDNENK 66  
DB 394 NTSNTATSNNTNNNN-----NNSSSGSEKRRKKNNNNNNGCPAVLLAARDKEI----- 442  
QY 67 PDKCALKEIKVQIROIKEOEKAAANID-----EVKSDVSTGGGVIDKDALGPMML 120  
DB 443 -----KALLDELQRLRAEOGTHLVQIQRLEHLEVKRHHIRL-EARLDKQOINBALA 494

```

QY 121 FALDGEFFVNLGVNVFVSENVTOYLRYNOEELMKNKSVSYLHVGDTHEFYKNLLPKSI 180
DB 495 EA-----TALSAASATNNNNNSOSSDNKK--LNTFAERMOASSNAD-----LPSS- 539
QY 181 VNGSGSEPRRRNSHTNCRLVLRP---DSEEGH-----DNODANOKYETMOCFAVS 233
DB 540 -----TKRPAEEDDEEDQAMLVDESEADKPED-----S 571
QY 234 OPKSKEGEDLOS-----CLICVARVRPKERVLPSSSEFTTRDLOGKITSL 283
DB 572 HHDDDEDEDEBEAVNATTTTNSNEIKKEQHSPLDLVNLSPNSAIAAAAAAAAAAACAN 631
QY 284 DSTMRA-----AMKPG-----WEDLYRRCIOKRFHAGESESVYAKRH----- 322
DB 632 DPNKFQALLIBRTKALAEALKNASDALSEDAHHQOQHQQOHHQOHHQOHHQOHHQO 691
QY 333 HHEVLRQGLASFQYRBSLQGTLYAOTKSKLIRSOTTNE-----POLVI 368
DB 692 HHHHLOQ-----OPNSGNSNPASNNHHGHNLGHGLHPSSAH 731
QY 369 SLHMLHRQNYCYMNPDLTGOTMGRPLNPISNSPAHALCSGNGODMTLSSNINPFIN 428
DB 732 HLHQTTESSNSSTPTAAGNNNGS--NNSSNT-----NANSTAOQLAASLASTLN 780
QY 429 GPKEDMGMRGFGSGGMNHVSGMOATTPQGSNYALKNPSPOSSPGMNFQOPTSMLESP 488
DB 781 GTSKLM-----QEDSNGLAAYA-----MAHAHQAAALGPFGLGRFQFRA 823
QY 489 RHRMSPGVA-GSPRIIPS--QFSPAGSLHSPVVC-----SSTGNSHYTN 531
DB 824 AQVAAGGDRGHYRFADSELDLPFGASMAGRLESLPKDPMKAELOEMLRYNMCKYAN 883
QY 532 SSUAL-----QALSEBGHVSIGSSLASP---DLKMGNLQNSPYNNMPPRLSKGSDS 582
DB 884 QALDTLHSRVRRELSVH--NIGQRFKAKYILGISQGTV--SELLSKRPWKDKTEGKR 939
QY 583 KDCFGLYGEPEEGTGAQESSCHPEQEKETNDPILPRAVSSERADGQRLDHSKOTKIL 642
DB 940 DSYKMHMAMACDDNAVMLKSLIP-----KKDSGLROYAG--RGAGGAGGDSMEDETA 992
QY 643 QLLTTKSDQMEPSPLASL-----SDTNKDSGLSPGSGSTHGTSLKEKNKILH 691
DB 993 HILSEASSILMKOSSVAQHREQERSHGEGEDSHSNEDEKS--PQSGCTSPFEKVENOLKH 1050
QY 692 RLDDSSSPVOLAKITLAVATGKDLDSQESSSTAPSEVITIKQEPVPRKKENALLRYLIDK 751
DB 1051 QHLNPEQAAAOQEREREQEREEOQRLRHDDQDMARLYOELIARPRRETAFPSEFL-- 1107
QY 752 DDTKDIGLPEITPKLERLDSKTDPAASNKLIAKTEKEEMSFEPDQDQSGELEDNLEETID 811
DB 1108 -----FSPSLFGGAACMPGASNAFPMADENNRHFE-----RELAKIQOHO 1151
QY 812 DLQNSQLPQLPPTPRGAPAGSVDKQAIINDL--MQLTAENSPVTYVGAOKTALRISOST 869
DB 1152 QOQAQAQAOAPPN-----FSSLALAQOQVNLGAODLSLAAAAAKDKIKLQSSSLEHSAGS 1207
QY 870 FNNRPGOLGRLPNONPLDITL--OSPAGGEPPIRMS-----SPYSVIPIQGM 919
DB 1208 SSCGKDEBRDAYPS-----SLHGKKSBEGGTRPAPAPPSGPGTGAGAPPTAAPTG- 1259
QY 920 MGNQMGIMQGNLQNSGTGIMGNSASRPTMPS--GEMAPQSSAVRVTCOAATTSAMNRVQ 977
DB 1260 -----GASSN--SAAPSPLSNLSILRPALSSQGEFFAATASPLQRMASINSLSLTQRPV 1310
QY 978 GGMTRNPAASLTPMRPSSO--PGQROTQLOSVNMIGPSEL--EMNNGGQOYSCQQAAPPNOT 1033
DB 1311 -----TPHHSRPPRPTKAVLVRPITQOQFDMFNNTATEDIVRRVKEALSQYISIQ----- 1359
QY 1034 APWPESTLPIQASFAQONRQPFSSPPDLLCPHPAASPSDEGALLDQL-----YLALRN 1089
DB 1360 RLFGBSVYIGLSQ-----GSVSDLLARPKP-----WHMLTQKREPIIRKMK 1400

```

```

QY 1090 EDGLEIDRALGIPELVVSQAVDEDEQFSODSNTIMLEQKARVFPQOYASQAOMAGSYS 1149
DB 1401 F--LEDEN-----AVHKLASQYKIAPEKILMRGS-----YSGSPQMPQGLASKWQAAS--L 1448
QY 1150 PMQDPNFTMGORBSYATILRMQPRGLRPTGIVQNPOLRLQLOLHRLQAOQNRPLMNO 1209
DB 1449 PMQKMSSELKLOEPRQAOHLMQOQMAAMSAMQOO--QVQAQOQAOQAOQAOQOHLQOO 1506
QY 1210 ISVNSVNLTLRPGVTPAPIAQMIAQORERILMNHILRORMHQOQVOQORTLMMRQGS 1269
DB 1507 -----AQOHLQOO-----QHLAQOO--HPHQOHHQ----- 1529
QY 1270 LNMTPSNVAPSGMATMSNPRIPOANAOQFPPRYNGISQOQDPDFTT-GATTP-----OSP 1324
DB 1530 -----AAAAAALHHQSMILNPSGLRPPQHA--ISLPPGAGGAQPGCGNGSSNPSSEKKP 1585
QY 1325 LMSPP-----RMAHT--OSPMMQOSQANPAYOAPSDIN----- 1354
DB 1586 MLMPHGTNAMSRLQIHMSPTVYEEMALTOIDLDTITTKIKEALLANNIGOKIRGEAVL 1645
QY 1355 GMAQGNMGNSMFSQOQPPHF-----GQOANTSMY-----SNMNMNINVMATITGMSMN 1405
DB 1646 GLSQGSV--SELLSKRPKMHMLSTKGRPFTIRMQLMSLDANNVERLOLTKERRRARRR 1703
QY 1406 QMTG-----QISMITSVTSTVSTGSLSSMGPEQY 1432
DB 1704 RSTGPNQODNSDITSNDTNDFTYSSPQPGSV 1735

```

```

RESULT 36
TRX_DROVI STANDARD: PRT: 3828 AA.
AC Q24742:
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Trithorax protein.
GN
OS Drosophila virilis (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_Taxid=7244;
RN [1]
RP
RP MEDLINE=96100387; PubMed=8555104;
RA Tiliid S., Sedkov Y., Mizrokh L., Mazo A.;
RT "Conservation of structure and expression of the trithorax gene
between Drosophila virilis and Drosophila melanogaster.";
RL Mech. Dev. 53:113-122(1995).
CC -!- FUNCTION: FUNCTIONS IN SEGMENT DETERMINATION THROUGH INTERACTION
WITH GENES OF TRITHORAX (BX-C) AND ANTENNAPEDIA (ANT-X) COMPLEXES.
CC IF CAN BEHAVE AS AN ACTIVATOR OF BX-C.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: BELONGS TO THE TRANSCRIPTION FACTOR TRITHORAX FAMILY.
CC -!- SIMILARITY: CONTAINS 1 'SET' DOMAIN.
CC -!- SIMILARITY: CONTAINS 5 PHD-TYPE ZINC FINGERS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
DB EMBL; Z50038; CA90349.1; -.
DB HSSP; P19793; 2NUL.
DB Flybase; FBgn0014844; Dvir\trx.
DB InterPro; IPR003889; Flyrich_C.
DB InterPro; IPR003888; Flyrich_N.
DB InterPro; IPR001965; PHD.
DB InterPro; IPR003616; PostSET.

```





```

Db 177 QGT-----HIIQSLDGLAAVAADGRLYISVYIIGLSQVEMTSSITFDYHOA 228
QY 167 DHEFVANKLLPKSIVNGSGNSG-----EPPRRNSHTFNCRL 203
Db 229 DHEF-IADQIGLISLTSGGGGGSSSSGGGGGAGGAGMASPTSGASDGGSTHCTNNPDV 287
QY 204 VKPLPDSSEEGHNOF-----AHQYE-----TWOC---FAVSOPKS 237
Db 288 AASTQTQASTGKYKGYDRSFCVRMKSTLTCLKRCKHSSGYSRASDATSCNNGNNAKN 347
QY 238 IKKEGELQGLICVARRV-----PKMERPV-----PS-----SESFTT 272
Db 348 VKNPGSNYSVYLCKLRPOYTSHRKSGPPLGMYALALPPEVHEIRECDMEVT 407
QY 273 RODLOGKITS-----LDTSTWRAAMKPGMEDLVRCIOKF-HAOHEGESVYAKRH 323
Db 408 RVNNDLRAVANCEPRVSDLLDYS-----EDLVNKSILYSLCHA---EDANRLRKSH 454
QY 324 HEVLRQGLAFSQYRFSLSQGLVAOYTKSKLIRSQTTNEPQVYLSHML-HREQ--- 377
Db 455 SDLEKQVLTLYYRLNKKSGGYTWTQTCATVCSKTKNADEQNICVNYVISNRENNMI 514
QY 378 -NVCVAMPDLTGQTMKRLPISNSPAHOA-----LCSGPGQDMTLSSINPPIGPK 431
Db 515 LDCQQLRP-----SPDSIKHEGGLGNDKSSSGPGGD----- 545
QY 432 EOMGMPMGREGSGS-GMNHVSGMOATTPQGSNTYALKMNSQSOSPQGNPCPTSMLSPRH 490
Db 546 -----ASGEKSHLSAGD-----MKLNSPTDEG----- 570
QY 491 RMSPFVAGSPRIIPPSQFSPAGSLHSPVGCSSGTGNSHSYTNSSINALQALSEGH---GVS 547
Db 571 -----HSHRGKRSAAASH---GSSMNSLTMLKIDSPPTPGVE 604
QY 548 LGSSLASPDLMGNLQNSPVNMNPPRLSKMGSLSKOCFGLYGPSPGCTTGOAASSCH- 605
Db 605 IDSCVLPPTVA-----TPPAATPPVQ-----STKRKRKTASOAE 641
QY 606 -PGOKETDNNPLPRAVSSERAGOSRLHDSKQTKLLQTLTTKSDOMERSPLASLSDT 664
Db 642 DQGEQYISQPLPKPLTMMQROQPR-----SRLPSTV-----DEQSSNADS----- 685
QY 665 NKDSTGSLPGSGSTHGTSLEKHKILHRLQDSSPYDLAKLAETATGKDISQSSSTAP 724
Db 686 -----AVKDLEQAMGSKHLPSPAVAVSVAPRNTDFESADSLKQOQ--- 724
QY 725 GSEVTIKOEFPSPKKKENALLRYLLDDDKDGLPEITPKLERLSDKTDPA----- 776
Db 725 -----OQOQUDPWEKSSSTI-----OWIGTPYQQPPAPY-----PATALLRQLY 762
QY 777 SNTKLIAKTEKEEMSFEPGCDQPSSELDNLEIILQNSQLPOLFPDTRGAPAGSVDK 836
Db 763 ANRSVRIRARARQPT---GVGPG-----VFYGDQOTGRLP-----TPPGS-----E 801
QY 837 QAITNDLMQJLAENSPLYTGAQKTAIRISOSTFNNRPGQGLGLPNQNLPLDITLQSP 896
Db 802 SSYENQYLOLHSAASGCHP-GGOKT---SADAFTNLVSTYGGY---HSSIDYHNAMATPP 853
QY 897 TGAGPPPIRINSSPYVILPQGMNGMGNMGNOGNSLGSMTGIMSASRPTMPSGMAP 956
Db 854 SSVSP---RDSN-----QEGKA--PVLASNGGYDAPPLGQYAT---SSGDVVP 897
QY 957 QSSAVRVCATTSAMNRPVOGGMIRNPASIPMRPSSOPGQROTLOSOVNIGPSELEM 1016
Db 898 -----ATLPLKPAQ-----SYATMHPSSGSTT 919
QY 1017 NMGPQYTSQOQAP 1029
Db 920 TEGGVYTSNLDQF 932

```

```

CBP_HUMAN
ID CBP_HUMAN STANDARD: PRT; 2442 AA.
AC Q92793; Q16376; 000147;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE 01-MAR-2002 (Rel. 41, Last annotation update)
DE CREB-binding protein.
GN CREBBP OR CBP.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97385172; PubMed=9238046;
RA Sobulo O.M., Borrow J., Tomek R., Reshmi S., Harden A.,
RA Schlegelberger B., Housman D., Doggett N.A., Rowley J.D.,
RA Zelenik-Je N.J.;
RT "MLL is fused to CBP, a histone acetyltransferase, in therapy-related
RT acute myeloid leukemia with a t(11;16)(q23;p13.3).";
RT Proc. Natl. Acad. Sci. U.S.A. 94:8732-8737(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97321049; PubMed=9177780;
RA Giles R.H., Petrij F., Dauwerse H.G., den Hollander A.I.,
RA Lushnikova T., van Ommen G.J.B., Goodman R.H., Deaven L.L.,
RA Doggett N.A., Peters D.J.M., Breuning M.H.;
RT "Construction of a 1.2-kb contig surrounding, and molecular analysis
RT of, the human CREB-binding protein (CBP/CREBBP) gene on chromosome
RT 16p13.3.";
RL Genomics 42:96-144(1997).
RN [3]
RP SEQUENCE OF 1-405 FROM N.A.
RX MEDLINE=96376968; PubMed=8782817;
RA Borrow J., Stanton V.P., Andresen J.M., Becher R., Behn F.G.,
RA Chaganti R.S.K., Civin C.I., Distcheke C., Dube I., Frischauf A.M.,
RA Horstman D., Mitelman F., Volinia S., Watmore A.E., Housman D.E.;
RT "The translocation t(8;16)(p11;p13) of acute myeloid leukemia fuses
RT a putative acetyltransferase to the CREB-binding protein.";
RL Nat. Genet. 14:33-41(1996).
CC -!- FUNCTION: MEDIATES CAMP-GENE REGULATION BY BINDING SPECIFICALLY TO
CC PHOSPHORYLATED CREB PROTEIN. ACTING AS A COACTIVATOR, CBP AUGMENTS
CC THE ACTIVITY OF PHOSPHORYLATED CREB TO ACTIVATE TRANSCRIPTION OF
CC CAMP-RESPONSIVE GENES.
CC -!- SUBUNIT: INTERACTS WITH SMAD1, SMAD2 AND SMAD3.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- DISEASE: INVOLVED IN ACUTE LEUKEMIAS BY CHROMOSOMAL TRANSLOCATIONS
CC T(8;16)(P11;P13) INVOLVING CBP AND MOZ, AND T(11;16)(Q23;P13.3)
CC INVOLVING CBP AND MLL.
CC -!- DISEASE: DEFECTS IN CREBBP ARE THE CAUSE OF RUBINSTEIN-TAYBI
CC SYNDROME (RSTS), A DISORDER CHARACTERIZED BY CRANIOFACIAL
CC ABNORMALITIES, BROAD THUMBS, BROAD BIG TOES, MENTAL RETARDATION
CC AND A PROPENSITY FOR DEVELOPMENT OF MALIGNANCIES.
CC -!- SIMILARITY: CONTAINS 1 BROMODOMAIN.
CC -!- SIMILARITY: CONTAINS 1 ZZ-TYPE ZINC FINGER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U47741; AAC51770.1; -
DR EMBL: U85962; AAC51331.1; -
DR EMBL: U89354; AAC51339.1; -
DR EMBL: U89355; AAC51340.1; -
DR MIM: 600140; -
DR InterPro: IPR001487; Bromodomain.
DR InterPro: IPR003101; KIX.
DR InterPro: IPR000197; TAZ_finger.
DR InterPro: IPR000433; ZnF_ZZ.

```

Query Match	3.2%	Score 243;	DB 1;	Length 2442;
Best Local Similarity	22.0%;	Pred. No. 0.0014;		
Matches 161;	Conservative 72;	Mismatches 232;	Indels 268;	Gaps 31;

QY	80/	EELIDDDQNSQL-----	-----	POLFEDHRRPGAPACSVYKOAIIINDLMOLVAENSPV	853		
Dd	1853	QOIOHRLOQOQLMRBRMATWTRNTENVQOSILPSTSAPEPTPOCP-----	-SIPCTPQ	1904			
OY	854	TPVGAKTARLISOSTFNNPNRPGQLRLPLNQLPDITLIQSPTCAGEPPIRNSSPV	913				
Dd	1905	PPAOPQSPPVSMSBAPFS-----	-VARTQP-----	PTTYSTGKTYSQVAPPAPPAOPPPAA	1955		
OY	914	IPOGMNGMOGITNOGNIGNSGTGMIGNASRPMTMPSGEAPPOSSAVRVTCATTSAMN	973				
Dd	1956	VEAROJTEREAO-----	OOOHLYRVN-----	INNS-----	MPCR-----	TGMGPQSOM	1995
OY	974	RPOVGGINRRPASISMPRESSOPCQRQTLOSOVMNIGPBELMNNGRGOYSQQQAAPROT	1033				
Dd	1996	APVSLNPFRNOVSGVPMSMPGO-----	-----	WOQAPLPQO	2029		
OY	1034	APWPESILLPIDOASFASONRPGSSPDLLCPHPAESPSDEGALLDOLYLALRNFDGL	1093				
Dd	2030	QPMGGLPRPV-----	ISMQAANVAFRMPRSVQPPRSIPS-----	AL	2067		
OY	1094	EELDRALGITELYSQOQADVDPEDFSSODBSINIMLEOKAPVFPOQVASQAOAQSGSYSPMD	1153				
Dd	2068	QDDLRTLKTSPPSQOQOQV-----	INTIKSN-----	POLMAAFTKORTAKY-----	2108		
OY	1154	PNFTMGQRPSVATLRMQRPGRGLRPTGLVONORPOLRLQLOHRLQAOQONRPLMNQISTNV	1213				
Dd	2109	-----VANOPG-----	MQPOPGL-----	QSQPG-----	MQPQGMHQOQSL	2139	
OY	1214	SNNVL-----	TLRGCVPO-----	APINNQAOLAORELTN	1244		
Dd	2140	QNLANMAOGVPRREVPPROQOQAGMLNPGQOALNINMGNPNMASMNPOYREMLRJOLLO	2199				
OY	1245	QHLROROMOHOQVOQOTFLM-----	MRGGLNMTPSMVAPSGMTYM-----	1286			

RESULT	39
HRX_MOUSE	
ID	HRX_MOUSE
AC	P55200;
DT	01-OCT-1996 (Rel. 34, Created)
DT	01-OCT-1996 (Rel. 34, Last sequence update)
DT	01-MAR-2002 (Rel. 41, Last annotation update)
DE	zinc finger protein HRX (ALL-1) (Frgment) .
GN	MLL OR HRX OR ALL1.
OS	Mus musculus (Mouse) .
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus
OX	NCBI_TaxID=10090;

```

RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC STRAIN-C57BL/6J, AND B6/CBA; TISSUE=Spleen, and Lung;
RX MEDLINE=93317679; PubMed=8327517;
RA Ma O., Alder H., Nelson K.K., Chatterjee D., Gu Y., Nakamura T.,
RA Canaan E., Croce C.M., Stracusa L.D., Buchberg A.M.;
RT "Analysis of the murine All-1 gene reveals conserved domains with
RT human All-1 and identifies a motif shared with DNA
RT methyltransferases.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:6350-6354(1993).
CC - FUNCTION: POSSIBLY ACTS AS A TRANSCRIPTIONAL REGULATORY FACTOR.
CC MA REGULATE GENES INVOLVED IN SKELETAL FORMATION DURING
CC EMBRYONIC DEVELOPMENT.
CC - SUBCELLULAR LOCATION: Nuclear (By similarity).
CC - ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC - SIMILARITY: BELONGS TO THE TRANSCRIPTION FACTOR TRITHORAX FAMILY.
CC - SIMILARITY: CONTAINS 1 BROMODOMAIN.
CC - SIMILARITY: CONTAINS 1 SET DOMAIN.
CC - SIMILARITY: CONTAINS 3 PHD-TYPE ZINC FINGERS.
CC - SIMILARITY: CONTAINS 1 CXXC-TYPE ZINC FINGER.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL, L17069; AAA62593.1; -.
DR MGD; MGI:96995; M11.
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR003889; Fyric_C.
DR InterPro; IPR003888; Fyric_N.
DR InterPro; IPR001965; PHD.
DR InterPro; IPR003616; PostSET.
DR InterPro; IPR000093; Recr.
DR InterPro; IPR001214; SET.
DR InterPro; IPR002857; Znf-CXXC.
DR Pfam; PF00628; PHD; 3.

```





```

Db 3050 LSVPHQHLSFPAQAQSSFPNISPGLIGVPPPDQLGSEANQRTDITTTVAT 3109
Oy 1321 POSPLASPRMA--HTQSPMWOQSOANPAYQAPSDI-----NGMAQGN 1360
Db 3110 PSSGKAKRPISRLHTRKKKALPSSAPSVIAPSDVSNMTLINFPSQISNHPSLDGS 3169
Oy 1361 MGNLS-----MFSQOS-----PPHFGQOANTSMYSNNMINVSMATN---- 1397
Db 3170 LNPSSHRTVPNTRKSKSGIMVFEDAPLLPQSVGTATAGSSSTISQDTSHLISGPVS 3229
Oy 1398 --TCGMSMNMONTGOISMTSVTSVSTSG 1423
Db 3230 ALASGSSVLNVSMQTTAAPTSTSTVPG 3257

RESULT 40
TRX_DROME STANDARD; PRT; 3726 AA.
ID TRX_DROME STANDARD; PRT; 3726 AA.
AC P20659; Q27255; Q27327;
DT 01-FEB-1991 (Rel. 17, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Trithorax protein.
GN Trx.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
CX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90192757; PubMed=2107543;
RA Mazo A.M., Huang D.-H., Mozer B.A., David I.B.;
RT "The trithorax gene, a trans-acting regulator of the bithorax complex
in Drosophila, encodes a protein with zinc-binding domains."
RL Proc. Natl. Acad. Sci. U.S.A. 87:2112-2116(1990).
RN [2]
RP SEQUENCE FROM N.A., ALTERNATIVE SPLICING, AND CHARACTERIZATION.
RX MEDLINE=95009521; PubMed=7924966;
RA Sedkov Y., Tiliib S., Mizrokh L., Mazo A.;
RT "The bithorax complex is regulated by trithorax earlier during
Drosophila embryogenesis than is the Antennapedia complex, correlating
with a bithorax-like expression pattern of distinct early trithorax
transcripts."
RL Development 120:1907-1917(1994).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=OREGON-R;
RA Tiliib S., Sedkov Y., Mizrokh L., Mazo A.;
RT "Conservation of structure and expression of the trithorax gene
between Drosophila virilis and Drosophila melanogaster."
RL Mech. Dev. 53:113-122(1995).
RN [4]
RP CHARACTERIZATION.
RX MEDLINE=95047388; PubMed=7958911;
RA Kuzin B., Tiliib S., Sedkov Y., Mizrokh L., Mazo A.;
RT "The Drosophila trithorax gene encodes a chromosomal protein and
directly regulates the region-specific homeotic gene fork head."
RL Genes Dev. 8:2478-2490(1994).
CC -1- FUNCTION: FUNCTIONS IN SEGMENT DETERMINATION THROUGH INTERACTION
WITH GENES OF BITHORAX (BX-C) AND ANTENNAPEDIA (ANT-X) COMPLEXES.
CC IT CAN BEHAVE AS AN ACTIVATOR OF BX-C.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- MISCELLANEOUS: THIS PROTEIN HAS BEEN EXPERIMENTALLY SHOWN TO BIND
ZINC.
CC -1- SIMILARITY: BELONGS TO THE TRANSCRIPTION FACTOR TRITHORAX FAMILY.
CC -1- SIMILARITY: CONTAINS 1 'SET' DOMAIN.
CC -1- SIMILARITY: CONTAINS 5 PHD-TYPE ZINC FINGERS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by, and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; M31617; AAA29025.1; -
DR EMBL; Z50152; CAA90514.1; -
DR EMBL; Z50152; CAA90513.1; -
DR EMBL; Z31725; CAA83516.1; -
DR EMBL; Z31725; CAA83515.1; -
DR PIR; A35085; A35085.
DR HSSP; P19793; 2NLL.
DR TRNSPAC; T00850; -
DR FlyBase; FBgn0003862; trx.
DR InterPro; IPR003889; Fyricl_C.
DR InterPro; IPR003888; Fyricl_N.
DR InterPro; IPR001965; PHD.
DR InterPro; IPR003616; PostSET.
DR InterPro; IPR001214; SET.
DR InterPro; IPR001841; Znf_ring.
DR Pfam; PF00628; PHD; 2.
DR Pfam; PF00856; SET; 1.
DR SMART; SM00542; FYRC; 1.
DR SMART; SM00541; FYRN; 1.
DR SMART; SM00249; PHD; 4.
DR SMART; SM00308; PostSET; 1.
DR SMART; SM00184; RING; 3.
DR SMART; SM00317; SET; 1.
DR PROSITE; PS0280; SET; 1.
KW Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;
KW Nuclear protein; Developmental protein; Activator;
KW Alternative splicing.
FT ZN_FING 1266 1347 PHD-TYPE 1.
FT ZN_FING 1348 1393 PHD-TYPE 2.
FT ZN_FING 1421 1482 PHD-TYPE 3.
FT ZN_FING 1734 1793 PHD-TYPE 4 (ATYPICAL).
FT ZN_FING 1794 1844 PHD-TYPE 5 (ATYPICAL).
FT DOMAIN 3599 3708 SET.
FT DOMAIN 512 516 POLY-SER.
FT DOMAIN 565 570 POLY-ASP.
FT DOMAIN 661 664 POLY-SER.
FT DOMAIN 905 910 POLY-SER.
FT DOMAIN 1576 1582 POLY-GLN.
FT DOMAIN 2298 3027 GLN-RICH.
FT DOMAIN 3032 3040 POLY-SER.
FT DOMAIN 3181 3184 POLY-GLN.
FT DOMAIN 3220 3225 POLY-GLU.
FT VARSPPLIC 1 368 MISSING (IN SHORT ISOFORM).
FT CONFLICT 2025 P -> PWLTSPKFLGLSTHGGLLMLLGVVRLKGG
(IN REF. 1).
FT CONFLICT 2341 R -> S (IN REF. 1).
FT CONFLICT 2392 G -> S (IN REF. 1).
SQ SEQUENCE 3726 AA; 400575 MW; D2756E50763D1CF5 CRC64;

Query Match 3.2%; Score 241.5; DB 1; Length 3726;
Best Local Similarity 18.7%; Pred. No. 0.0027;
Matches 294; Conservative 204; Mismatches 596; Indels 477; Gaps 69;

Oy 108 GYIDDALCPMLLEALDGFPPVNE--GNVPVSENVOTLKRNOEELM--MNSYSI 162
Db 1890 GSEVRQGLGATVPRESDSYEAVVPINFLCSRLYSSKPEKIVETVTTIONSSSTLTA 1949
Oy 163 LHVQ-----DHT-----EFVKNLIPRSIVGSGMSSEPPRR 193
Db 1950 LDVGKRYNYVDHNPNSKEVQGLMAOIARHNTSLASELEN-----GCTWSEFPNP 2002
Oy 194 NSHTNCRMLVKPLPDSEEGHNOE-----AHQKYMOCFAV-----S 233
Db 2003 NS-----CVPDPONTFEEEPQOQADLLPPELKDAIFEDLPHELDGISMLDIFYDXT 2055

```



This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**

**THIS PAGE BLANK (USPTO)**